

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:21:33 ; Search time 24.9726 Seconds

(without alignments)
5623.200 Million cell updates/sec

Title: US-08-353-485-2

Perfect score: 2641

Sequence: 1 MRKLSLFSLAVALLSLLCWG.....QLNLTSVAGQVKVLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 2641 | 100.0 | 497 | 2 AAR96025 | Aar96025 P. gingiv |
| 2 | 2641 | 100.0 | 497 | 2 AAW69483 | Aaw69483 Haemagglu |
| 3 | 2641 | 100.0 | 2628 | 2 AAR96030 | Aar96030 P. gingiv |
| 4 | 2641 | 100.0 | 2628 | 2 AAW69488 | Aaw69488 Haemagglu |
| 5 | 951.5 | 36.0 | 1706 | 2 AAW24786 | Aaw24786 Prtr anti |
| 6 | 944.5 | 35.8 | 1704 | 2 AAR70188 | Aar70188 Arg-gingi |
| 7 | 944.5 | 35.8 | 1704 | 2 AAW34843 | Aaw34843 Arg-gingi |
| 8 | 944.5 | 35.8 | 1704 | 3 AAY67396 | Aay67396 Arg-gingi |
| 9 | 944.5 | 35.8 | 1704 | 4 AAU08938 | Aau08938 P. gingiv |
| 10 | 943.5 | 35.7 | 1687 | 2 AAR96033 | Aar96033 P. gingiv |
| 11 | 943.5 | 35.7 | 1687 | 2 AAW69495 | Aaw69495 Haemagglu |
| 12 | 816 | 30.9 | 1358 | 2 AAR96032 | Aar96032 P. gingiv |
| 13 | 816 | 30.9 | 1358 | 2 AAW69494 | Aaw69494 Haemagglu |
| 14 | 812.5 | 30.8 | 1732 | 2 AAR96029 | Aar96029 P. gingiv |
| 15 | 812.5 | 30.8 | 1732 | 2 AAW24787 | Aaw24787 PrtX anti |
| 16 | 812.5 | 30.8 | 1732 | 2 AAW69487 | Aaw69487 Haemagglu |
| 17 | 806.5 | 30.5 | 1087 | 2 AAR96028 | Aar96028 P. gingiv |
| 18 | 806.5 | 30.5 | 1087 | 2 AAW69486 | Aaw69486 Haemagglu |
| 19 | 791.5 | 30.0 | 419 | 4 AAU03572 | Aau03572 P. gingiv |
| 20 | 662.5 | 25.1 | 419 | 4 AAU03574 | Aau03574 P. gingiv |
| 21 | 546.5 | 20.7 | 450 | 2 AAR96021 | Aar96021 P. gingiv |
| 22 | 546.5 | 20.7 | 450 | 2 AAW69489 | Aaw69489 Haemagglu |
| 23 | 546.5 | 20.7 | 456 | 2 AAR96023 | Aar96023 P. gingiv |
| 24 | 546.5 | 20.7 | 456 | 2 AAR96022 | Aar96022 P. gingiv |
| 25 | 546.5 | 20.7 | 456 | 2 AAW69491 | Aaw69491 Haemagglu |

ALIGNMENTS

RESULT 1

AAR96025
ID AAR96025 standard; protein; 497 AA.

XX AC AAR96025;

XX AC

DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX

DE P. gingivalis haemagglutinin hgaA.

XX

KW Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.

XX

OS Porphyromonas gingivalis; strain 381.

XX

PN WO9617936-A2.

XX

PD 13-JUN-1996.

XX

PE 11-DEC-1995; 95WO-US016108.

XX

PR 09-DEC-1994; 94US-00353485.

XX

PA (UYFL) UNIV FLORIDA.

XX

PA (UABR-) UAB RES FOUND.

XX

PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX

DR WPI; 1996-287181/29.

XX

DR N-PSDB; AAT30649.

XX

PT Porphyromonas gingivalis genes and proteins - used in the detection and

XX

PT vaccination against periodontal disease.

XX

PS Claim 5; Page 47-51; 153pp; English.

XX

CC A portion (AAR96025) of P. gingivalis 381 haemagglutinin hgaA (see also

XX

CC AAR96030) was identified as the product of clone ST2 (AAT30649) isolated

XX

CC from a P. gingivalis 318 genomic library. The haemagglutinin can be obtd.

XX

CC animals against periodontal disease. Expression in Salmonella cells

XX

CC allows prodn. of a live vaccine. The haemagglutinin can also be used to

XX

CC detect the presence of anti-P. gingivalis antibodies and to raise

XX

CC monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-2003 to

XX

SQ Sequence 497 AA;

Query Match 100.0%; Score 2641; DB 2; Length 497;
 Best Local Similarity 100.0%; Pred. No. 2.5e-194;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKNSLSFLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60
 DB 1 MRKNSLSFLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60

QY 61 ARIILEAHVDWEDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120
 DB 61 ARIILEAHVDWEDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180

QY 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVVLNESFDTQ 240
 DB 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVVLNESFDTQ 240

QY 241 TLPNGWTMIDADGDNHMLSTINVTNTATHTGCGAMFSKSWTASGAKIDLSPDNYLVTVP 300
 DB 241 TLPNGWTMIDADGDNHMLSTINVTNTATHTGCGAMFSKSWTASGAKIDLSPDNYLVTVP 300

QY 301 KVTVPENGKLSYVWSSQVPWNEHYGVFLSTTGNEAANFTIKLLEETGSKDKPAPMNLVK 360
 DB 301 KVTVPENGKLSYVWSSQVPWNEHYGVFLSTTGNEAANFTIKLLEETGSKDKPAPMNLVK 360

QY 361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
 DB 361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

QY 421 YRDNVVIQAQNLAAATTFNQENAVPGQYNYCVVKYTAGVSPKVKCDVTVEGSNEFAHVQNL 480
 DB 421 YRDNVVIQAQNLAAATTFNQENAVPGQYNYCVVKYTAGVSPKVKCDVTVEGSNEFAHVQNL 480

QY 481 TGSVAGQKVTLLKWDAPN 497
 DB 481 TGSVAGQKVTLLKWDAPN 497

RESULT 2

AAW69483

ID AAW69483 standard; protein; 497 AA.

XX AAW69483;

AC AAW69483;

DT 22-DEC-1998 (first entry)

XX Haemagglutinin protein hgaA.

XX Haemagglutinin protein; periodontal disease; vaccine; hgaA.

XX Porphyromonas gingivalis.

OS US824791-A.

XX US824791-A.

PD 20-OCT-1998.

XX 11-DEC-1995; 95US-00570311.

XX 08-SEP-1988; 88US-00241640.

PR 25-JAN-1991; 91US-00647119.

PR 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulskie-Fox A, Lepine G;

XX WPI; 1998-582627/49.

DR N-PSDB; AAV58870.

XX

PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 protease poly:peptide(s)).

XX Claim 1; Col 37-44; 101pp; English.

XX This sequence is encoded by a Porphyromonas gingivalis gene of the
 invention. This sequence represents the hgaA haemagglutinin protein. The
 polypeptides are used to produce antibodies to organisms associated with
 periodontal disease. The antibodies are also used in purification and
 identification procedures. The genes and polypeptides are used as
 vaccines against periodontal disease

XX Sequence 497 AA;

Query Match 100.0%; Score 2641; DB 2; Length 497;
 Best Local Similarity 100.0%; Pred. No. 2.5e-194;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKNSLSFLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60
 DB 1 MRKNSLSFLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60

QY 61 ARIILEAHVDWEDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120
 DB 61 ARIILEAHVDWEDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180

QY 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVVLNESFDTQ 240
 DB 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVVLNESFDTQ 240

QY 241 TLPNGWTMIDADGDNHMLSTINVTNTATHTGCGAMFSKSWTASGAKIDLSPDNYLVTVP 300
 DB 241 TLPNGWTMIDADGDNHMLSTINVTNTATHTGCGAMFSKSWTASGAKIDLSPDNYLVTVP 300

QY 301 KVTVPENGKLSYVWSSQVPWNEHYGVFLSTTGNEAANFTIKLLEETGSKDKPAPMNLVK 360
 DB 301 KVTVPENGKLSYVWSSQVPWNEHYGVFLSTTGNEAANFTIKLLEETGSKDKPAPMNLVK 360

QY 361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420
 DB 361 SEGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDDVAVSGEGSSNDYTYTV 420

QY 421 YRDNVVIQAQNLAAATTFNQENAVPGQYNYCVVKYTAGVSPKVKCDVTVEGSNEFAHVQNL 480
 DB 421 YRDNVVIQAQNLAAATTFNQENAVPGQYNYCVVKYTAGVSPKVKCDVTVEGSNEFAHVQNL 480

QY 481 TGSVAGQKVTLLKWDAPN 497
 DB 481 TGSVAGQKVTLLKWDAPN 497

RESULT 3

AAR96030

ID AAR96030 standard; protein; 2628 AA.

XX AAR96030;

AC AAR96030;

XX 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX P. gingivalis haemagglutinin hgaA.

DE Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.

XX Porphyromonas gingivalis; strain 381.

OS

XX Key

FH Location/Qualifiers

FT Peptide

5..21

/label= sig_peptide

FT

XX WO9617936-A2.
 PN
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 PR 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Progulsk-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX WPI; 1996-287181/29.
 DR N-PSDB; AAT30654.
 XX
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 PT
 XX Claim 6; Page 93-101; 153pp; English.
 PS
 XX P. gingivalis 381 haemagglutinin hgaA (AAR96030) was identified as the
 CC product of the hga gene (AAT30654) isolated as an EcoRV fragment of
 CC genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
 CC 24), can be obtd. from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
 CC also be used to detect the presence of anti-P. gingivalis antibodies and
 CC to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
 CC 2003 to standardise OS field)
 XX
 XX SQ Sequence 2628 AA;
 Query Match 100.0%; Score 2641; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 2.9e-193;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRKNSLFSVLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVXDLRDPDPAGM 60
 DB 1 MRKNSLFSVLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVXDLRDPDPAGM 60
 QY 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFHWANGTIPAGLYDPFFYKVPVNA 120
 DB 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFHWANGTIPAGLYDPFFYKVPVNA 120
 QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VQRQGGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVILNESFDTQ 240
 DB 181 VQRQGGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVILNESFDTQ 240
 QY 241 TLNGWMTIDADGNGHNLSTINVTATHTGDMFSGKSWTASGGAKIDLSPDNLYLTP 300
 DB 241 TLNGWMTIDADGNGHNLSTINVTATHTGDMFSGKSWTASGGAKIDLSPDNLYLTP 300
 QY 301 KVTVPENGKLSYVWSSQVPTNBNHYGVFLSTTGTNEAANFTIKLLEETLGSCKPAPMNLVK 360
 DB 301 KVTVPENGKLSYVWSSQVPTNBNHYGVFLSTTGTNEAANFTIKLLEETLGSCKPAPMNLVK 360
 QY 361 SEGKLPAPYQERTIDLSAVAGQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
 DB 361 SEGKLPAPYQERTIDLSAVAGQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
 QY 421 YRDNVVIAQNLAAATTFNQENVAPQYNYCVVEKYTAGVSPKCKVDVTVEGSNEFAHQNL 480
 DB 421 YRDNVVIAQNLAAATTFNQENVAPQYNYCVVEKYTAGVSPKCKVDVTVEGSNEFAHQNL 480
 QY 481 TGSVAGQKVTWKWDAPN 497
 DB 481 TGSVAGQKVTWKWDAPN 497

RESULT 4

AAW69488 standard; protein; 2628 AA.
 XX
 AC AAW69488;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hgaA.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN US5824791-A.
 XX
 PD 20-OCT-1998.
 XX
 XX 11-DEC-1995; 95US-00570311.
 XX
 PR 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulsk-Fox A, Lepine G;
 XX WPI; 1998-582627/49.
 DR N-PSDB; AAV58875.
 XX
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease poly.peptide(s).
 PT
 XX Claim 1; Col 91-110; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 XX SQ Sequence 2628 AA;
 Query Match 100.0%; Score 2641; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 2.9e-193;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRKNSLFSVLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVXDLRDPDPAGM 60
 DB 1 MRKNSLFSVLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVXDLRDPDPAGM 60
 QY 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFHWANGTIPAGLYDPFFYKVPVNA 120
 DB 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFHWANGTIPAGLYDPFFYKVPVNA 120
 QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VQRQGGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVILNESFDTQ 240
 DB 181 VQRQGGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVILNESFDTQ 240
 QY 241 TLNGWMTIDADGNGHNLSTINVTATHTGDMFSGKSWTASGGAKIDLSPDNLYLTP 300
 DB 241 TLNGWMTIDADGNGHNLSTINVTATHTGDMFSGKSWTASGGAKIDLSPDNLYLTP 300
 QY 301 KVTVPENGKLSYVWSSQVPTNBNHYGVFLSTTGTNEAANFTIKLLEETLGSCKPAPMNLVK 360

```

Db 301 KVTVPENGLSYWSSVQPTNEHYGVFLSTGTGNEAANFIKLEELTSGDKPAPMLVX 360
QY 361 SEGKLPAPYQERTIDLSAVAGQOVYLAFRHFNSTGIFRLYLDVAVSGEGSNDYTYTV 420
Db 361 SEGKLPAPYQERTIDLSAVAGQOVYLAFRHFNSTGIFRLYLDVAVSGEGSNDYTYTV 420
QY 421 YRDNVIAQNLAAATFQENAVPQYNYCDEVKYTAGVSPKVCCKDVTEGSEFAHVONL 480
Db 421 YRDNVIAQNLAAATFQENAVPQYNYCDEVKYTAGVSPKVCCKDVTEGSEFAHVONL 480
QY 481 TGSVAGQKVLTKWDAPN 497
Db 481 TGSVAGQKVLTKWDAPN 497

RESULT 5
AAW24786
ID AAW24786 standard; protein; 1706 AA.
XX
AC AAW24786;
XX
DT 17-OCT-2003 (revised)
DT 25-NOV-1997 (first entry)
XX
DE PrrR antigenic protein complex.
XX
KW Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW PrrR; PrrR45; PrrR15; PrrR17; PrrR27; haemagglutinin; adhesin;
KW therapy; diagnosis; vaccine; antigen.
XX
OS Porphyromonas gingivalis; strain W50.
XX
FH Key
FH Peptide 1..227 Location/Qualifiers
FT /label= Pro-pro_peptide
FT Cleavage-site 227..228
FT Protein 228..719
FT /label= PrrR45
FT /note= "45 kDa Arg-specific thiol protease"
FT 719..720
FT Cleavage-site 720..1138
FT Protein /label= PrrR44
FT /note= "44 kDa adhesin"
FT 1138..1139
FT Cleavage-site 1139..1273
FT Protein /label= PrrR15
FT /note= "15 kDa adhesin"
FT 1273..1274
FT Cleavage-site 1274..1431
FT Protein /label= PrrR17
FT /note= "17 kDa adhesin"
FT 1431..1432
FT Cleavage-site 1432..1706
FT Protein /label= PrrR27
FT /note= "27 kDa adhesin"
XX
PN W09716542-A1.
XX
PD 09-MAY-1997.
XX
PF 30-OCT-1996; 96WO-AU000673.
XX
PR 30-OCT-1995; 95AU-00006275.
XX
PA (UYME ) UNIV MELBOURNE.
PA (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI Reynolds EC, Bhogal PS, Slakeski N;
XX
DR WPI; 1997-272112/24.
DR N-PSDB; AAT78850.
XX

```

```

PT New antigenic protein complex from Porphyromonas gingivalis - comprising
PT Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
PT prevention and treatment of periodontal disease.
XX
XX Example 1; Fig 8b; 68pp; English.
XX
CC A PrrR-PrrK cell surface protein of Porphyromonas ginivalis (PG)
CC comprises a 300 kDa complex composed a 45 kDa arginine-specific thiol
CC protease and 44, 15, 17 and 27 kDa adhesins encoded by the prrR gene
CC (AAT78850), and a 148 kDa lysine-specific thiol protease and 39, 15 and
CC 44 kDa adhesins (see AAW24787) encoded by the prk gene (AAT78851). A
CC claimed antigenic complex comprises at least one multimeric protein
CC complex of PrrR and PrrK each containing at least one adhesin domain, the
CC complex having a mol. wt. of over 200 kDa, and preferably comprises all 9
CC proteins of the PrrR-PrrK complex (see also AAW24780-85). It can be used
CC in a claimed composition to elicit an immune response directed against
CC PG, and in a claimed method of reducing the prospect of PG infection
CC and/or severity of disease. Antibodies directed against the complex are
CC claimed for use in treating PG infection. Unlike whole PG cells or other
CC previously prepared antigens based on fimbriae or the capsule, the PrrR-
CC PrrK complex or component parts are safe and effective antigens. (Updated
CC on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1706 AA;
XX
Query Match 36.0%; Score 951.5; DB 2; Length 1706;
Best Local Similarity 31.9%; Pred. No. 1.3e-63;
Matches 237; Conservative 74; Mismatches 156; Indels 275; Gaps 20;
QY 22 TAAAGQ-----GPKTAPSVTHQAVQK--GIR---TSKVQDLRDPFAGMARILIEAH 68
Db 671 TATTQGGKVLTKWDAPSTKTNTATNTARSVDGIRLVLVSDAPELLRSGQAEIVLEAH 730
QY 69 DVWEDGTGYQWLWDADHNOYGASIPBE--SFWFANGTIPAGLYDPPRYKVPVADAFSP 126
Db 731 DVWMDGSGYQILLDADHDQYGVIPSDTHLW--PNCVSPANLFAFFYTPENADPCSP 789
QY 127 TNFVLDGTASADIPAGTYDYVIINPNPGI--IYIVGEGVSGKNDYVVEAGKYTHFTVQROG 185
Db 790 TNWIMDGTASVNIPIAGTYDFAIAAPQANAKIWAGQGTKEDDYVFPAGKYHFLMKMG 849
QY 186 PGDAASVV----- 193
Db 850 SGDTELTISGEGGSDYTYTYRDTGKIKEGLTATTPEDGDTAGNHEYCVVKYTAGVS 909
QY 194 -----VTGEGNEFAVQNLQWSVSGOTVTLTQAPAS-----DKRTYVINE 235
Db 910 PKVCKDVTVESNEFAVPVQNLTGSAVGQKVTLKWDAPNPTNPNNPNNPNNPNTTISE 969
QY 236 SFDQTPLPNGWTMIDADGDGHNMLSTINVTANTHTGDGAMFSGKSWTASGAKIDLSPDN 295
Db 970 SFE-NGIPASWKTIADGDGHG--KPCNAPGIAGYNSNGCVYSESF--GLGGIGV-LTPDN 1025
QY 296 YLVTPTKVTVPENGLKSYWSSQ--VPWTNEHYGVFLSTGTGNEAANFTIKLEETL---GSD 351
Db 1026 YLITPALDLPLNGGKLTFWCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1085
QY 352 KPAPM----- 356
Db 1086 SPEAMRGRIQGTWRQKTVDLIPAGTKYVAFRHFQSTDMFYDLDEVEIKANGKRADFTET 1145
QY 357 -----NLVKS----- 361
Db 1146 ESSTHGEAPAEWTTIDADGGQGWLCISGQLDWLTAHGGTINVVVSFWNGMALNPONYL 1205
QY 362 ----- 361
Db 1206 ISKDVGTATKVKYXYAVNDGFPFGDHYAVMISKTGNAGDEFTVVFEEPTNGINKGARFGL 1265
QY 362 ----EGVKLPAPYQERTIDLSAVAGQOVYLAFRHFNSTGIFRLYLDV--AVSGEGSND 415
Db 1266 STEADGAKPOSWTIERIVDLP--AGTK-YVAFRHYNCSDNLNLIILDDIQTIMGSPSTPTD 1322

```

QY 416 YTVTVEDNVVIAQNLAAATTENQENVAPGOYNYCVKVTAGYSPKVKCDVTVEGSENEA 475
 Db 1323 YTVTVEDNVVIAQNLAAATTENQENVAPGOYNYCVKVTAGYSPKVKCDVTVEGSENEA 475

QY 476 HVQNLGTSAYGQKVTLLKWDAPN 497
 Db 1382 PVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 6
 ID AAR70188 standard; protein; 1704 AA.
 XX AAR70188;
 AC
 DT 25-MAR-2003 (revised)
 DT 21-SEP-1995 (first entry)
 XX
 XX Arg-gingipain-2 prepolyprotein.
 XX
 KW Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
 KW arginine-specific protease.
 XX
 OS Porphyromonas gingivalis.
 XX
 PH Key Location/Qualifiers
 FT Protein 228..719
 FT /label= Protease
 FT /note= "corresponds to Arg-gingipain-1"
 FT 720..1091
 FT /label= Hemagglutinin
 FT 1092..1429
 FT /label= Hemagglutinin
 FT 1430..1704
 FT /label= Hemagglutinin
 FT
 XX WO9507286-A1.
 XX
 XX 16-MAR-1995.
 XX
 XX 09-SEP-1994; 94WO-US010283.
 XX
 XX 10-SEP-1993; 93US-00119361.
 PR 21-OCT-1993; 93US-00141324.
 PR 24-JUN-1994; 94US-00265441.
 XX
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA
 XX
 PI Travis J, Potempa J, Barr PJ, Pavloff N;
 XX
 XX WPI; 1995-123373/16.
 DR N-PSDB; AAQ83489.
 XX
 XX DNA encoding Arg-gingipain proteins - used to develop prods. for
 PT detection, treatment and prevention of periodontal disease.
 PT
 XX
 XX Disclosure; Page 70-77; 89pp; English.
 PS
 XX
 CC A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
 CC 2) were isolated from P. gingivalis strains H66 (ARCC 33277) and W50
 CC (ATCC 53973). The sequences of the proteins were used to design PCR
 CC primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
 CC libraries were screened with a probe based on amino acids 11-22 of the AG
 CC protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
 CC is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 CC
 XX Sequence 1704 AA;
 SQ

Query Match 35.8%; Score 944.5; DB 2; Length 1704;
 Best Local Similarity 31.9%; Pred. No. 4.4e-63;
 Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TARAQG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPDPAGMARIIIEAH 68
 Db 671 TATQGGKVTWKDASTKTATNTARSVDGRELVLVSDAPELRGQRAIVLEAH 730

QY 69 DWEDGTGYQMLWDADHNOYGASIPBE--SFWFANGTIPAGLYDPPFFYKVPVNADASFSP 126
 Db 731 DVWDSGGYQILLDADHDQYGVIPSDTHLW--PNCSPANLFAPEYTPVNADPSCSP 789

QY 127 TNFVLDTASADIPAGTYDYIINPAGI--IYIVGEGVSKGNDYVVEAGKTYHTVORQG 185
 Db 790 TNNIMDGTASVNI PAGTYDFAIAAQAQANAKINIAQGQPTKEDDYVFEAGKKYHFLMKMG 849

QY 186 PGDAASVW-----
 Db 850 SGDTELTISEGGSDYTYTVYRDGTIKIEGLTATTFEEDGVATGNHEYCVKVTAGVS 909

QY 194 -----VTGEGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----DKRTVVLNESF 237
 Db 910 PKVKCDVTVEGSENEFAPVQNLGTSAGVQKVTLLKWDAPNGTPNPNPNPNPGTTLSESF 969

QY 238 DTOTLPNGWTMIDADGHNWLTSTINVTATHTGDGAMESKSWTASGGAKIDLSPDNYL 297
 Db 970 E-NGIPASWKTIDADGDGHW--KPGNAPGIAGYNSGCVYSESF--GLGG-GV-LTPDNYL 1025

QY 298 VTPKVTVPENGKLSYVVSQ--VPWTNEHYGVFLSTTGNAAAFIKLLEETLGS----- 350
 Db 1026 ITPALDLPNGGKLTFFWVCAQADANYASEHYAVASSIGNDASNETNALLEETITAKGVRS 1085

QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
 Db 1086 EAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADETTFES 1145

QY 367 ----PAPYQERTID-----LSAYAGQOV----- 385
 Db 1146 STHGEAPAEWTTIDADGCGQWMLCLSSGQDLTLAHGGTVNVASFWSNMGALNPDNYLIS 1205

QY 386 ----- 385
 Db 1206 KDVGTATKVKYVAVNDGFFGDHVAVMISKTGTNAGDFTVVFEETPNKNGGARFGLST 1265

QY 386 -----YLAPRHPNSTGIFRLYLDLV--AVSGGSSNDYTYTV 420
 Db 1266 EANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLAILDDIQFTMGGSPTPTDYTYTV 1325

QY 421 YRDNVIAQNLAAATTENQENVAPGOYNYCVKVTAGYSPKVKCDVTVEGSENEFAHYQNL 480
 Db 1326 YRDTKIIEGLTETTFEEDGVATGNHEYCVKVTAGYSPKCVNVTI--NPTQPNPVKNL 1384

QY 481 TGSAGVQKVTLLKWDAPN 497
 Db 1385 KAOPDGGDDVVLKWEAPS 1401

RESULT 7
 ID AAW34843 standard; protein; 1704 AA.
 XX AAW34843;
 AC
 XX 03-JUN-1998 (first entry)
 DT
 XX Arg-gingipain high molecular weight prepolyprotein sequence.
 DE
 XX Arg-specific gingipain protease; gingivalis; periodontal disease;
 KW vaccine; infection.
 KW
 XX Porphyromonas gingivalis.
 OS
 XX
 PH Key Location/Qualifiers
 FT Protein 1..227
 FT /note= "precursor protein"
 XX
 XX WO9734629-A1.

CC modulate Arg-gingipain proteinase activity inhibitors

XX Sequence 1704 AA;

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
Best Local Similarity 31.9%; Pred. No. 4.4e-63;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSVKDLRDPPIAGMARIIIEAH 68
Db 671 TATQOGKVTLKWDAPSTKTNATNTARSVDGRELVLVSDAPELLRSQAEIVLEAH 730
QY 69 DVWEDGTGYQMLWDADHNOYCASIPBE--SFWFANGTIPAGLYDPFFYKVPVNADASFSP 126
Db 731 DVWNGSGYQILLDADHDQGVQVTPSDTHLM--PNCVFNALFAPFYTTPENADPCSP 789
QY 127 TNFVLDGTASADIPAGTYDYVIINPNFI--IYIVGEGVSKNDYVVEAGKYHTFTVQROG 185
Db 790 TNMIMDGTASVNIPTAGTYDFAIAAPQANAKIWIAGQGTREDDYVFEAGKYHFLMKMG 849
QY 186 PGDAASVV----- 193
Db 850 SGDTEITISEGGSDTYTYVYRDGKIKEGLTATTEEDGVAATGNHEYCEVVKYTAGVS 909
QY 194 -----VTGEGNEFAFVQNLQWVSGQVTLTWQAPAS-----DKRTYVLNESF 237
Db 910 PKVKDVTVEGNEFAFVQNLGSAVQKVTLLKWDAPNGTPNPNPNPGTITLSESP 969
QY 238 DTQTLPNGTWIDADGSHNWLSTINYNTATHTGDMGAMFSKSWTAGGAKIDLSPNYL 297
Db 970 E-NGIPASWTIDADGSHGW--KPGNAPGIAGVNSGCVSESP--GLGGIGV-LTPDNYL 1025
QY 298 VTPKVTVPENKLSYVSSQ--VPMTNEHYGVFLSTTCNEANFTIKLLEETLGS----- 350
Db 1026 ITPALDLPNGKLTFFWCAQDANTASEHYAVYASSTGNDASNFTNALLETTITAKVRSP 1085
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 1086 EAIRGRIQGTWROKTVLDLPAGTKYVAFHFQSTDMFYDILDEVEIKANGKEADTFEFES 1145
QY 367 -----PAPYQERTID-----LSAYAGQQV----- 385
Db 1146 STHGEAPAEWTTIDADGGQGLCLSSGQDLMTAHGNTVVASFSWNGMALNPDNYLIS 1205
QY 386 ----- 385
Db 1206 KDTGATKVKYVAVNDGFFGDHVAVMISKGTGNAGDFTVVFETPNKNGGARFGLST 1265
QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db 1266 EANGAKPQSVMIERTVLDLPAGTKYVAFHYNCSLDNVLILLDDIQTWGGSPPTDYTYTV 1325
QY 421 YRDNVVIAQNLATTFNQENVAPQYNYCEVVKYTAGVSKVKDVTVEGNEFAHVQNL 480
Db 1326 YRDGTKIKELGTETTFEDGVTAGNHEYCEVVKYTAGVSKPCVNVNTI-NPTQPNPKVL 1384
QY 481 TGSVAVGQKVTLKWDAPN 497
Db 1385 KAQPDGDDVVLKWEAPS 1401

RESULT 9

AAU08938

ID AAU08938 standard; protein; 1704 AA.

XX

AC AAU08938;

XX

DT 18-DEC-2001 (first entry)

XX

DE P. gingivalis high molecular weight Arg-gingipain-2.

XX

XW Periodontitis; antiinflammatory; Arg-gingipain-2; AG-2; immunogen.

XX

OS Porphyromonas gingivalis.

XX Key Location/Qualifiers

FT Peptide 1..227

FT /label= Prepro_peptide

FT Region 599..619

FT /note= "Region of homology with cysteine proteases"

FT Region 670..674

FT /label= Proteolytic_component

FT Cleavage-site 719

FT Protein 720..1091

FT /label= HGP_44kDa

FT /note= "Haemagglutinin protein component"

FT Cleavage-site 1091

FT Protein 1092..1429

FT /label= HGP_17kDa

FT /note= "Haemagglutinin protein component"

FT Cleavage-site 1429

FT Protein 1430..1704

FT /label= HGP_27kDa

FT /note= "Haemagglutinin protein component"

XX US6274718-B1.

XX

XX 14-AUG-2001.

XX 25-JAN-2000; 2000US-00490931.

XX 10-SEP-1993; 93US-00119361.

XX 24-JUN-1994; 94US-00265441.

XX 09-SEP-1994; 94WO-US010283.

XX 08-NOV-1994; 94US-00336308.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX

XX Travis J, Potempa JS, Barr PJ, Pavloff N;

XX MPI; 2001-588904/66.

XX N-PSDB; AAS15242.

XX

XX New recombinant DNA molecule which encodes high molecular weight (mature) Arg-gingipain protein, useful for immunization against inflammation and tissue damage, comprises enzymatically active protease component and hemagglutinin component.

XX Claim 1; Col 29-41; 56pp; English.

XX

XX The invention relates to a recombinant DNA molecule encoding high molecular weight (mature) Arg-gingipain (AG) protein, which has an enzymatically active protease component (AG-2) and a haemagglutinin component, from P. gingivalis. The nucleic acid is useful for producing mature Arg-gingipain protein. Immunogenic compositions comprising Arg-gingipain are useful for immunising animals including humans against inflammatory response and tissue damage caused by an archaebacterium Porphyromonas gingivalis, which causes progressive periodontitis. Arg-gingipain is also useful for identifying agents that modulate Arg-gingipain proteinase activity, whether by acting on the proteinase itself or preventing the interaction of the proteinase with the protein in the gingival area, such as complement factors C3 or C5. The present sequence is Arg-gingipain-2

XX Sequence 1704 AA;

Query Match 35.8%; Score 944.5; DB 4; Length 1704;

Best Local Similarity 31.9%; Pred. No. 4.4e-63;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSVKDLRDPPIAGMARIIIEAH 68

Db 671 TATQOGKVTLKWDAPSTKTNATNTARSVDGRELVLVSDAPELLRSQAEIVLEAH 730

QY 69 DVWEDGTGYQMLWDADHNOYCASIPBE--SFWFANGTIPAGLYDPFFYKVPVNADASFSP 126

| | | | |
|-----------|---|--|------|
| Db | 731 | DVNDGSGYQILLDADHDQGVIPSDTHTLW-PNCSSVPANLFAPEYTVPENADPSCSP | 789 |
| Qy | 127 | TNFVLDTGASADI PACTYDVYLIINPNPGI-IYVGEVSKGNDYVVEAGKTHFTVQROG | 185 |
| Db | 790 | TNIMDGTASVNPAGTYDFAAQAQANAKIWIAGQGTKEDDYVFEAGKCYHFLMKMG | 849 |
| Qy | 186 | PGDAASV----- | 193 |
| Db | 850 | SGDGTSLTISEGGSDYTYTVYRDGTHKIKEGLTATTFEEDSGVATGNHEVCVEVKTAGVS | 909 |
| Qy | 194 | -----VTGEGNEFAPVQNLQWSVSGQTVTLTWOAPAS-----DKRTYVLNBSF | 237 |
| Db | 910 | PKVCKDVTVEGSEFAPVQNLGSAVGQVTLKWDAPNGTPNPENPNPNTTILSESF | 969 |
| Qy | 238 | DTQTLPNGWMTIDADGDGHNWLTSTINVTYATHTGDGAMFSKSWTASGAKIDLSPDNYL | 297 |
| Db | 970 | E-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSSEF-GLGGIGV-LTPDNYL | 1025 |
| Qy | 298 | VTPKVTVPENKLSYVWSQ-VPWTHNEHYGVFLSTTGNEAANFTIKLLEETLGS----- | 350 |
| Db | 1026 | ITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNTNALLLEETITAKGVRSF | 1085 |
| Qy | 351 | -----DKPAPMNLV-----KSEGVKL----- | 366 |
| Db | 1086 | FAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF | 1145 |
| Qy | 367 | -----PAPYQERTID-----LSAVAGQOV----- | 385 |
| Db | 1146 | STHGEPAPWTTIDADGDGQGWCLSSGQDLWLTAGGTVNVVASFWSNGMALNPONYL | 1205 |
| Qy | 386 | ----- | 385 |
| Db | 1206 | KDVTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTVVFETPENGKGGARFGLST | 1265 |
| Qy | 386 | -----YLAFRHFNSTGIFRLYLDV--AVSGESSNDYTYTV | 420 |
| Db | 1266 | EANGAKPOSWIERTVDLPAGTKYVAFRHYNSDLNILLDDIQFTMGSGPTPTDYTV | 1325 |
| Qy | 421 | YRDNVIAQNLAAATFNQENVAPGVNYCDEVKYTAGVSPKCKDVTVEGSENEFAHVQNL | 480 |
| Db | 1326 | YDGTGKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVNVTI-NPTQFNPVKNL | 1384 |
| Qy | 481 | TGSAVGQKVTLLKWDAPN | 497 |
| Db | 1385 | KAQPDGQGVVLKWEAPS | 1401 |
| RESULT 10 | | | |
| AAR96033 | | | |
| ID | AAR96033 standard; protein; 1687 AA. | | |
| XX | AAR96033; | | |
| AC | 16-OCT-2003 (revised) | | |
| DT | 04-SEP-1996 (first entry) | | |
| XX | P. gingivalis haemagglutinin hage. | | |
| DE | Haemagglutinin; hage; periodontal disease; vaccine; antibody. | | |
| KW | Porphyromonas gingivalis; strain FDC381. | | |
| OS | WO9617936-A2. | | |
| XX | 13-JUN-1996. | | |
| PD | 11-DEC-1995; 95WO-US016108. | | |
| PF | 09-DEC-1994; 94US-00353485. | | |
| XX | (UYFL) UNIV FLORIDA. | | |
| PA | (UABR-) UAB RES FOUND. | | |
| XX | | | |

| | | | |
|--|------|--|------|
| PI | 731 | DVNDGSGYQILLDADHDQGVIPSDTHTLW-PNCSSVPANLFAPEYTVPENADPSCSP | 789 |
| XX | 127 | TNFVLDTGASADI PACTYDVYLIINPNPGI-IYVGEVSKGNDYVVEAGKTHFTVQROG | 185 |
| DR | 790 | TNIMDGTASVNPAGTYDFAAQAQANAKIWIAGQGTKEDDYVFEAGKCYHFLMKMG | 849 |
| XX | 186 | PGDAASV----- | 193 |
| PT | 850 | SGDGTSLTISEGGSDYTYTVYRDGTHKIKEGLTATTFEEDSGVATGNHEVCVEVKTAGVS | 909 |
| XX | 194 | -----VTGEGNEFAPVQNLQWSVSGQTVTLTWOAPAS-----DKRTYVLNBSF | 237 |
| CC | 910 | PKVCKDVTVEGSEFAPVQNLGSAVGQVTLKWDAPNGTPNPENPNPNTTILSESF | 969 |
| CC | 238 | DTQTLPNGWMTIDADGDGHNWLTSTINVTYATHTGDGAMFSKSWTASGAKIDLSPDNYL | 297 |
| CC | 970 | E-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSSEF-GLGGIGV-LTPDNYL | 1025 |
| CC | 298 | VTPKVTVPENKLSYVWSQ-VPWTHNEHYGVFLSTTGNEAANFTIKLLEETLGS----- | 350 |
| CC | 1026 | ITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNTNALLLEETITAKGVRSF | 1085 |
| CC | 351 | -----DKPAPMNLV-----KSEGVKL----- | 366 |
| CC | 1086 | FAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF | 1145 |
| CC | 367 | -----PAPYQERTID-----LSAVAGQOV----- | 385 |
| CC | 1146 | STHGEPAPWTTIDADGDGQGWCLSSGQDLWLTAGGTVNVVASFWSNGMALNPONYL | 1205 |
| CC | 386 | ----- | 385 |
| CC | 1206 | KDVTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTVVFETPENGKGGARFGLST | 1265 |
| CC | 386 | -----YLAFRHFNSTGIFRLYLDV--AVSGESSNDYTYTV | 420 |
| CC | 1266 | EANGAKPOSWIERTVDLPAGTKYVAFRHYNSDLNILLDDIQFTMGSGPTPTDYTV | 1325 |
| CC | 421 | YRDNVIAQNLAAATFNQENVAPGVNYCDEVKYTAGVSPKCKDVTVEGSENEFAHVQNL | 480 |
| CC | 1326 | YDGTGKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVNVTI-NPTQFNPVKNL | 1384 |
| CC | 481 | TGSAVGQKVTLLKWDAPN | 497 |
| CC | 1385 | KAQPDGQGVVLKWEAPS | 1401 |
| Query Match 35.7%; Score 943.5; DB 2; Length 1687; | | | |
| Best Local Similarity 31.8%; Pred. No. 5.2e-63; | | | |
| Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20; | | | |
| Qy | 22 | TAAAGQ-----GPKTAPSVTHQAVQK--GIR---TSKVLDLDPPIAGMARILLEAH | 68 |
| Db | 652 | TATQGGQKVTLLKWDAPSTKTNTATNTAKSVGIRLWLLSDAPELRSGQAEIVLEAH | 711 |
| Qy | 69 | DVWEDGTGYQLMDADHNQYQASIPSE--SEFWANGTIPAGLYDPFFEYKVPVNADASFSP | 126 |
| Db | 712 | DVWNGSGYQILLDADHDQGVIPSDTHTLW-PNCSSVPANLFAPEYTVPENADPSCSP | 770 |
| Qy | 127 | TNFVLDTGASADI PACTYDVYLIINPNPGI-IYVGEVSKGNDYVVEAGKTHFTVQROG | 185 |
| Db | 771 | TNIMDGTASVNPAGTYDFAAQAQANAKIWIAGQGTKEDDYVFEAGKCYHFLMKMG | 830 |
| Qy | 186 | PGDAASV----- | 193 |
| Db | 831 | SGDGTSLTISEGGSDYTYTVYRDGTHKIKEGLTATTFEEDGVATGNHEVCVEVKTAGVS | 890 |
| Qy | 194 | -----VTGEGNEFAPVQNLQWSVSGQTVTLTWOAPAS-----DKRTYVLN | 235 |
| Db | 891 | PKVCKDVTVEGSEFAPVQNLGSAVGQVTLKWDAPNGTPNPENPNPNTTILSE | 950 |
| Qy | 236 | SFDQTLPNGWMTIDADGDGHNWLTSTINVTYATHTGDGAMFSKSWTASGAKIDLSPDN | 295 |
| Db | 951 | SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSSEF-GLGGIGV-LTPDN | 1006 |
| Qy | 296 | YLVTPKVTVPENKLSYVWSQ-VPWTHNEHYGVFLSTTGNEAANFTIKLLEETLGS--- | 350 |
| Db | 1007 | YLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNTNALLLEETITAKGV | 1066 |
| Qy | 351 | -----DKPAPMNLV-----KSEGVKL----- | 366 |
| Db | 1067 | SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETF | 1126 |
| Qy | 367 | -----PAPYQERTID-----LSAVAGQOV----- | 385 |
| Db | 1127 | ESSTHGEPAPWTTIDADGDGQGWCLSSGQDLWLTAGGTVNVVASFWSNGMALNPONYL | 1186 |
| Qy | 386 | ----- | 385 |
| Db | 1187 | ISKDVTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTVVFETPENGKGGARFGL | 1246 |
| Qy | 386 | -----YLAFRHFNSTGIFRLYLDV--AVSGESSNDYTY | 418 |
| Db | 1247 | STEANGAKPOSWIERTVDLPAGTKYVAFRHYNSDLNILLDDIQFTMGSGPTPTDYTY | 1306 |
| Qy | 419 | TVYRDNVIAQNLAAATFNQENVAPGVNYCDEVKYTAGVSPKCKDVTVEGSENEFAHVQ | 478 |
| Db | 1307 | TVYRDGTHKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVNVTI-NPTQFNPVK | 1365 |

QY 479 NLTSAGVQKVKLWDAPN 497
 DB 1366 NLKAQPDGGDVVLKWEAPS 1384

RESULT 11
 AAW69495
 ID AAW69495 standard; protein; 1687 AA.
 XX AC
 XX AAW69495;
 XX
 DT 22-DEC-1998 (first entry)
 XX DE
 XX Haemagglutinin protein hage.
 XX KW
 XX Haemagglutinin protein; periodontal disease; vaccine; hage.
 XX OS
 XX Porphyromonas gingivalis.
 XX PN
 XX US824791-A.
 XX PD
 XX 20-OCT-1998.
 XX PF
 XX 11-DEC-1995; 95US-00570311.
 XX PR
 XX 08-SEP-1988; 88US-00241640.
 XX PR
 XX 25-JAN-1991; 91US-00647119.
 XX PR
 XX 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 XX PA
 XX (UABR-) UAB RES FOUND.
 XX
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
 DR WPI; 1998-582627/49.
 DR N-PSDB; AAV58881.
 XX
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 FT protease poly(peptide(s)).
 XX
 XX Claim 1; Col 167-182; 101pp; English.
 XX
 XX This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hage haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX SQ Sequence 1687 AA;

Query Match 35.7%; Score 943.5; DB 2; Length 1687;
 Best Local Similarity 31.8%; Pred. No. 5.2e-63;
 Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

QY 22 TAAAG-----GPKTASVTHQVOK--GIR---TSKVQDLRDPFAGMARILEAH 68
 DB 652 TATQGVKVLKWDAPSTKTATNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAH 711
 QY 69 DVWEDGTGYQMLWDADHNOYGASIPER--SFWFANGTIPAGLYDPPEYKVPVNADASFSP 126
 DB 712 DVWNGSGYQILLDADHDQGVQVPSDTHLM--PNCVSPANLFAPEYTVFENADPSCSP 770
 QY 127 TNFVLDGTASADIPAGHYDYVVIINPNFGI-IYIVGEGVSKGNDVYVEAGKTYHFTVQROG 185
 DB 771 TNMIMDGTASVNPAGTYDFALAPQANAKIWIAGQPTKEDDYVFEAGKKYHFLMKMG 830
 QY 186 PGDAASV-----
 DB 831 SGDGTETLTISEGGSDYTYWYRDGTXKKEGLTATTTFEEDGVATGNHEYCVKTYTAGVS 890
 QY 194 -----VTGGGNEFAFVQNLQWSGQTVTLTWOAPAS-----DKRTVILNE 235

891 PKVCKDVTVEGSNEFAFVQNLQWSGQTVTLKWDAPNGTNPNPNPNPNGTITLSE 950
 236 SFDQTLPNGWTMIDADGDHNLSTINVYNTATHTGDGAMFSKSWTASGAKIDLSPDN 295
 951 SFE-NGIPASWKTIDADGDHGM-KPGNAPGIAGYNSNGCVSSEF-GLGGIGV-LTFDN 1006
 296 YLVTPKVTPENGKLSYVWSQ-VPWTHNEHYGVFLSTGTGNEAANFTIKLLEETLGS----- 350
 1007 YLITPDLPLNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNTALLETITAKGVR 1066
 351 -----DKPAPNVLV-----KSEGVKL----- 366
 1067 SPEALRGRIQGTWQKTVDLPAQTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTTF 1126
 367 -----PAPYQERTID-----LSAYAGQGV----- 385
 1127 ESSTHGEAPAEWTTIDADGQGWCLCLSSGGQLDMLTAHGNTNVVASFWSNGMALNPDNYL 1186
 386 ----- 385
 1187 ISKDVTKATKVKYVAVNDGPPGDHYAVMISKGTGNAGDFTVVFETPNKNGKARFGL 1246
 386 -----YLAFRHFNSTGIFRLYLDV--AVSCGSGSNDVY 418
 1247 STEANGAKPQSVWLTERTVDLPAGTKYVAFRHYNCSDLYILLDDIQTMGGSPTPTDTY 1306
 419 TVYRDNVVIAQNLAAATFNQENVAPQYQYCVKYEYTAGVSPKCKDVTVEGSNEFAHVQ 478
 1307 TVYRDGTXKKEGLTETTFEEDGVATGNHEYCVKYEYTAGVSPKCVNVTI-NPTQFNPVK 1365
 479 NLTSAGVQKVKLWDAPN 497
 1366 NLKAQPDGGDVVLKWEAPS 1384

RESULT 12
 AAR96032
 ID AAR96032 standard; protein; 1358 AA.
 XX AC
 XX AAR96032;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis hagD haemagglutinin.
 XX
 KW Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
 XX
 OS Porphyromonas gingivalis; strain FDC381.
 XX
 PN WO9617936-A2.
 XX
 XX 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 PR 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 XX (UABR-) UAB RES FOUND.
 XX
 XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 DR WPI; 1996-287181/29.
 DR N-PSDB; AAT30655.
 XX
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 5; Page 125-129; 153pp; English.
 XX
 XX P. gingivalis 381 haemagglutinin hagD (AAR96032) was identified as the
 CC product of the second open reading frame of the hagD gene (AAT30655)
 CC

QY 351 -----DKPAPMNLV-----KSRGVKL----- 366
 Db 738 SPEAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSDMFYIDLDEVEIKANGKRADEFTEP 797
 QY 367 -----PAPVOERTID-----LSAYAGQV----- 385
 Db 798 ESSTHGEAPAEWTITDADGGQDWLCLSSGQLDLTAHGTTNVVAFSFWNGMALNPNDYL 857
 QY 386 ----- 385
 Db 858 ISKDVGTATKVKYVAVNDGFPGDHYAVMLSKTGNAGDTTVVFEETPNKNGGARFGL 917
 QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSSEGSNDYTY 418
 Db 918 STEANGAKPQSVWERTVLDLPAGTKYVAFRHYNCSDLDYILLDDIQFTMGSSPTPTDYTY 977
 QY 419 TVYRDNVIAQNLAAATFNQENAVPGQYVCEVKYTAGVSPKCKDVTVEGSNEFAHVQ 478
 Db 978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCEVKYTAGVSPKVCVNVVTI-NPTQFNPK 1036
 QY 479 NLGTSVAGQKVTLLKWDAPN 497
 Db 1037 NLKAQPDGDDVVLKWEAPS 1055
 RESULT 14
 AAR96029
 ID AAR96029 standard; protein; 1732 AA.
 AC AAR96029;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis porphyain.
 XX
 KW Porphyain; haemagglutinin; periodontal disease; vaccine; antibody.
 XX
 OS Porphyromonas gingivalis; strain W12.
 XX
 FH Key Location/Qualifiers
 FT Region 688..708
 FT /note= "Pro-Asn repeat region type 1"
 FT Region 887..952
 FT /note= "Pro-Asn repeat region type 2"
 FT Region 946..967
 FT /note= "Pro-Asn repeat region type 1"
 FT Region 985..1006
 FT /note= "Pro-Asn repeat region type 3"
 FT Region 1041..1100
 FT /note= "Pro-Asn repeat region type 4"
 FT Region 1341..1405
 FT /note= "Pro-Asn repeat region type 2"
 FT Region 1430..1451
 FT /note= "Pro-Asn repeat region type 3"
 FT Region 1488..1547
 FT /note= "Pro-Asn repeat region type 4"
 FT Region 1607..1650
 FT /note= "Pro-Asn repeat region type 2"
 XX WO9617936-A2.
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 PR 09-DEC-1994; 94US-00353485.
 XX
 PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 PI Proguliske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX

DR WPI; 1996-287181/29.
 XX N-PSDB; AAT30653.
 PT Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 PS Claim 5; Page 76-81; 153pp; English.
 XX
 CC P. gingivalis W12 cysteine protease, porphyain (AAR96029), was
 CC identified as the product of the prp gene (AAT30653) isolated from P.
 CC gingivalis W12 genomic DNA. The porphyain shows homology to the
 CC haemagglutinins (see also AAR96026-28 and AAR96030-33) of P. gingivalis
 CC 318. It can be obt. from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows prodn. of a live vaccine. The porphyain and
 CC haemagglutinins can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1732 AA;
 Query Match 30.8%; Score 812.5; DB 2; Length 1732;
 Best Local Similarity 30.0%; Pred. No. 6.4e-53;
 Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;
 QY 22 TAAAGGPKT---APSVTHQAVQKGIKRTSKVKDLRD-----PIPAGMARIILEA 67
 Db 693 TATTQKQKVTLKWEAPS-----AKKAGSREVKRIKIGDLFTIEPANDVRAEKVVLAA 747
 QY 68 HDVWEDGTQYMLWDADHNOYASIPESFWFANGTIAGLYDP-FEYKVPVNADAFSP 126
 Db 748 DNVMGDNTGYQLLDADHNTFGSVIPATPLF-TGTASSNLNYSANFEYLPANADPVVTT 806
 QY 127 TNFVLDGTASADIPACTYDVYIINPNP-GLIYVGEG---VSKGNDYVVEAGKTVHFTV 181
 Db 807 QNIIVTQGEVWIPGGVYDYCIITNPEPASGKMMTAGDGNOPARYDFTFEAGKXYTFTM 866
 QY 182 QROGPGDAASV----- 193
 Db 867 RRAGMGDTMEVEDDSPASYTVTVYRDGKIKESGLTATTFEEDGVAAGNHEYCEVKYT 926
 QY 194 -----VTGEGNEFAPVQNLQMSVSGQVTVLWQAP-----ASDKRTVVLN 234
 Db 927 AGVSPKCKDVTVVEGSNEFAPVQNLGTSSVQKVTLLKWDAPNGTPNPFPNGTTL 986
 QY 235 ESPDTQTLNPGWTMIDADGHNWLSITNVYNTATHTGDGAMFSKSWTASGAKIDLSPD 294
 Db 987 ESFE-NGIPASWKTIDADGGHGW-KPGNAPGIAGYNSNGCVSESP-GLGGIGV-LTPD 1042
 QY 295 NYLVTPKVTVPENGKLSYVVSQ-VPWTNEHYGVFLSTTGNAAANFTIKLLEETLGS--- 350
 Db 1043 NYLITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDASNTNALLEETITAKGV 1102
 QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
 Db 1103 RSPKAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSDMFYIDLDEVEIKANGKRADEFTE 1162
 QY 367 -----PAPVOERTID-----LSAYAGQV----- 385
 Db 1163 FESSTHGEAPAEWTITDADGGQDWLCLSSGQLDLTAHGTSNVVSVFWSNGMALNPNDY 1222
 QY 386 ----- 385
 Db 1223 LISKDVGTATKVKYVAVNDGFPGDHYAVMLSKTGNAGDTTVVFEETPNKNGGARFG 1282
 QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSSEGSNDYTY 417
 Db 1283 LSTEANGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDNLILLDDIQFTMGSSPTPTDYTY 1342
 QY 418 YTVYRDNVIAQNLAAATFNQENAVPGQYVCEVKYTAGVSPKCKDVTVEGSNEFAHV 477
 Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCEVKYTAGVSPKCKVDVTV-NSTQFNPV 1401

QY 478 QNLTGSAV--GQKVLKWDAP 496
 AAW24787
 Db 1402 QNLTRAEQAPNSMDAILKWNAP 1422

RESULT 15
 ID AAW24787 standard; protein; 1732 AA.
 XX AAW24787;
 AC AAW24787;
 XX 17-OCT-2003 (revised)
 DT 25-NOV-1997 (first entry)
 XX
 XX PrtK antigenic protein complex.
 DE
 XX Periodontal disease; cell surface protein; thiol protease; endopeptidase;
 KW PrtK; PrtK48; PrtK39; PrtK45; PrtK44; haemagglutinin; adhesin; therapy;
 KW diagnosis; vaccine; antigen.
 XX
 XX Porphyromonas gingivalis; strain W50.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..228
 FT /label= Pro-pro-peptide
 FT Cleavage-site 228..229
 FT Protein 229..737
 FT /label= PrtK48
 FT /note= "48 kDa Lys-specific thiol protease"
 FT Cleavage-site 737..738
 FT Protein 738..1156
 FT /label= PrtK39
 FT /note= "39 kDa adhesin"
 FT Cleavage-site 1156..1157
 FT Protein 1157..1291
 FT /label= PrtK15
 FT /note= "15 kDa adhesin"
 FT Cleavage-site 1291..1292
 FT Protein 1292..1732
 FT /label= PrtK44
 FT /note= "44 kDa adhesin"
 XX
 XX WO9716542-A1.
 PN
 XX 09-MAY-1997.
 PD
 XX
 XX 30-OCT-1996; 96WO-AU000673.
 PF
 XX
 XX 30-OCT-1995; 95AU-00006275.
 PR
 XX
 XX (UYNM) UNIV MELBOURNE.
 PA (VICT-) VICTORIAN DAIRY IND AUTHORITY.
 XX
 XX Reynolds EC, Bhogal PS, Slakeski N;
 PI
 XX WPI: 1997-272112/24.
 DR
 XX N-PSDB; AAT78851.
 DR
 XX
 XX New antigenic protein complex from Porphyromonas gingivalis - comprising
 PT Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
 PT prevention and treatment of periodontal disease.
 PT
 XX Example 1; Fig 9b; 69pp; English.
 PS
 XX
 XX A PrtR-PrtK cell surface protein of Porphyromonas gingivalis (PG)
 CC comprises a 300 kDa complex composed a 48 kDa lysine-specific thiol
 CC protease and 39, 15 and 44 kDa adhesins encoded by the prtK gene
 CC (AAT78851), and a 45 kDa arginine-specific thiol protease and 44, 15, 17
 CC and 27 kDa adhesins (see AAW24786) encoded by the prtR gene (AAT78850). A
 CC claimed antigenic complex comprises at least one multimeric protein
 CC complex of PrtR and PrtK each containing at least one adhesin domain, the
 CC complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
 CC proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used

CC in a claimed composition to elicit an immune response directed against
 CC PG, and in a claimed method of reducing the prospect of PG infection
 CC and/or severity of disease. Antibodies directed against the complex are
 CC claimed for use in treating PG infection. Unlike whole PG cells or other
 CC previously prepared antigens based on fimbriae or the capsule, the PrtR-
 CC PrtK complex or component parts are safe and effective antigens. (Updated
 CC on 17-OCT-2003 to standardise OS field)
 XX
 XX Sequence 1732 AA;
 SQ

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
 Best Local Similarity 30.0%; Pred. No. 6.4e-53;
 Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

QY 22 TAAAGGGPKT---APSVTHQAVQKGIKTSKVKDLRD-----PIPMAGRIILEA 67
 Db 693 TATTGGQKVLKWEAPS-----AKKAGSRVRKRGIGDLFTVIEPANDVRANEAKVILAA 747
 QY 68 HDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDP-FEYKVPVNDASFSF 126
 Db 748 DNVAGDNTGYQFLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVANADPVVTT 806
 QY 127 TNFVLDGTASADIPAGTYDYVVIINPNP-GIIYIVGEG---VSKGNDYVWEAGTYHETV 181
 Db 807 QNIIVTGGGEVVIIPGGVYDYCITNPEPASGRMWIAGDGGNQPARYDDFTFEAGKKTFTM 866
 QY 182 QROGFGDAASVV-----
 Db 867 RRAGMGDTMEVEDDSPASVYTVYRDGTVIKKGLTATTFEEDGVAAGNHEYCVVEKYT 926
 QY 194 -----VTGEGGNEFAFPQNLQMSVSGQTTLTWQAP-----ASDKRTVYVLN 234
 Db 927 AGVSPKCKDVTVGSGNEFAFPQNLTGSSVQKVLKWDAPENGTPNPNPNPFGTTLS 986
 QY 235 ESFDTQILPNCWTMIDADGDGHNWLSITNVNTATHGDCGAMFSKSWTASGAKIDLSPD 294
 Db 987 ESFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPD 1042
 QY 295 NYLVTPKVTPENGKLSYVWSSQ-VPWTNEHYGVFLSTTGNEAANFTIKLEETLGS--- 350
 Db 1043 NYLITPALDLPGGKLTFFWCAQDANYASEHYAYASSTGNDASNFTWALLEETITAGV 1102
 QY 351 -----DKPAPNVLV-----KSPGVKL----- 366
 Db 1103 RSPKAIRGRIQGTWRFQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAEDTET 1162
 QY 367 -----PAPYQERTID-----LSAYAGQGV----- 385
 Db 1163 FESSTHGEAPAEWTTIDADGGQGWLCSSGQLDWLTHAGGNSVVSFSSWNGMALNPDY 1222
 QY 386 -----
 Db 1223 LISKDVTKATKVKYVYVNDGFGPDGHYAVMISKTGNAGDFTVVFEETENGKNGARFG 1282
 QY 386 -----YLAFRHFSTGIFRLYLDV--AVSGESSNDYT 417
 Db 1283 LSTEANGAKPOSVMERTVDLPAGTKYVAFRHYNCSDNLNLTLLDDIQFTMGSSPTPTDYT 1342
 QY 418 YTVTRDNRVIAQNLAAATTFNCEAVPGQYNCVEVKYTAGVSPKVKCDVTVGEGNEFAHV 477
 Db 1343 YTVTRDGTIKKEGTEFTTFEEDGVATGNHEICVBEVKYTAGVSPKVKCDVTVV-NSTQFNVP 1401
 QY 478 QNLTGSAV--GQKVLKWDAP 496
 Db 1402 QNLTRAEQAPNSMDAILKWNAP 1422

Search completed: May 18, 2004, 11:32:40
 Job time : 29.9726 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:29:09 ; Search time 9.14177 Seconds
(without alignments)

5229.534 Million cell updates/sec

Title: US-08-353-485-2

Perfect score: 2641

Sequence: 1 MRKLSLFLSLAVLLSLLCWG.....QNLTGSAVGQKVTWKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pirl:*

2: pirl2:*

3: pirl3:*

4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | DB ID | Description |
|------------|-------|---------|--------|----------|---------------------------|
| 1 | 2641 | 100.0 | 2628 | 2 T28651 | hemagglutinin A - |
| 2 | 950.5 | 36.0 | 1526 | 2 A49763 | gingipain R (EC 3.1.11.1) |
| 3 | 944.5 | 35.8 | 1704 | 2 A55426 | gingipain R (EC 3.1.11.1) |
| 4 | 812.5 | 30.8 | 1732 | 2 T30836 | lysine-specific cy |
| 5 | 494.5 | 18.7 | 991 | 2 I40229 | arginyl endopeptid |
| 6 | 156 | 5.9 | 1231 | 1 A48490 | endo-1,4-beta-xyla |
| 7 | 153 | 5.8 | 1231 | 2 P90696 | hypothetical prote |
| 8 | 152 | 5.8 | 1234 | 2 S72640 | endo-1,4-beta-xyla |
| 9 | 150 | 5.7 | 1483 | 2 C97012 | probably celluloso |
| 10 | 147 | 5.6 | 1983 | 2 G86643 | hypothetical prote |
| 11 | 144.5 | 5.5 | 781 | 2 T36143 | probable secreted |
| 12 | 138.5 | 5.2 | 2468 | 2 A83412 | hypothetical prote |
| 13 | 137.5 | 5.2 | 635 | 2 S19011 | endo-1,4-beta-xyla |
| 14 | 137 | 5.2 | 2013 | 2 A11489 | probable peptidogl |
| 15 | 135.5 | 5.1 | 1433 | 1 A36734 | bacillopeptidase F |
| 16 | 135 | 5.1 | 1655 | 2 E97835 | hypothetical prote |
| 17 | 132 | 5.0 | 1651 | 2 JC1340 | outer membrane pro |
| 18 | 131.5 | 5.0 | 725 | 2 A90255 | hypothetical prote |
| 19 | 129.5 | 4.9 | 1873 | 2 T30944 | surface protein pr |
| 20 | 129 | 4.9 | 940 | 2 AB1744 | internalin protein |
| 21 | 129 | 4.9 | 1349 | 2 A11476 | cell surface prote |
| 22 | 128 | 4.8 | 1348 | 2 AH1115 | cell surface prote |
| 23 | 128 | 4.8 | 4936 | 2 AH2515 | hypothetical prote |
| 24 | 127.5 | 4.8 | 5188 | 2 B85547 | probable RTX famil |
| 25 | 127 | 4.8 | 2710 | 2 A37052 | toxin A - Clostrid |
| 26 | 126.5 | 4.8 | 908 | 2 AE2254 | hypothetical prote |
| 27 | 126.5 | 4.8 | 1148 | 2 S72635 | exo-poly-alpha-gal |
| 28 | 125.5 | 4.8 | 940 | 2 AD1374 | internalin protein |
| 29 | 125 | 4.7 | 1503 | 2 T18266 | cycloinulo-oligosa |

30 124.5 4.7 868 2 AF3204
31 124.5 4.7 875 2 H90371
32 124 4.7 1616 2 T17884
33 124 4.7 2013 2 AD1129
34 123.5 4.7 208 2 AF1718
35 122 4.6 729 2 T35028
36 122 4.6 1217 2 F97177
37 121.5 4.6 1530 2 AH1396
38 121 4.6 1142 2 C97080
39 120.5 4.6 587 2 A55368
40 120.5 4.6 1448 2 AI2007
41 120 4.5 364 1 JE0292
42 120 4.5 1939 2 D97316
43 119 4.5 639 2 A38368
44 119 4.5 2205 2 T08615
45 118.5 4.5 694 2 B84331

ALIGNMENTS

RESULT 1

T28651

hemagglutinin A - Porphyromonas gingivalis

C;Species: Porphyromonas gingivalis

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001

C;Accession: T28651

R;Han, N.; Whitlock, J.; Progulsk-Fox, A.

Infect. Immun. 64, 4000-4007, 1996

A;Title: The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381 contains four 1

A;Reference number: Z20494; MUID:97047672; PMID:8926061

A;Accession: T28651

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-2628 <HAN>

A;Cross-references: EMBL:U41807; NID:gl552410; PID:gi469916; PIDN:AAB17128.1

C;Genetics:

A;Gene: haga

Query Match 100.0%; Score 2641; DB 2; Length 2628;

Best Local Similarity 100.0%; Pred. No. 2.1e-163;

Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | | |
|----|-----|---|------|----|
| QY | 1 | MRKLSLFLSLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVOKGIRTSKVKDLRDP | PAGM | 60 |
| DB | 1 | MRKLSLFLSLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVOKGIRTSKVKDLRDP | PAGM | 60 |
| QY | 61 | ARTILEAHDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDPPFYKVPVNA | 120 | |
| DB | 61 | ARTILEAHDVWEDGTGYQMLWDADHNQYGASIPESFWFANGTIPAGLYDPPFYKVPVNA | 120 | |
| QY | 121 | DASFSPNFDGTASADIPAGTYDYVVIINPNPGIYIVGEGYKGNVYVVEAGKTYHFT | 180 | |
| DB | 121 | DASFSPNFDGTASADIPAGTYDYVVIINPNPGIYIVGEGYKGNVYVVEAGKTYHFT | 180 | |
| QY | 181 | VORQSGPDAAVVVTVGEGNEFAPVQNLQWSVSGQVTLTQWAPADSKRTYVLNESFDTC | 240 | |
| DB | 181 | VORQSGPDAAVVVTVGEGNEFAPVQNLQWSVSGQVTLTQWAPADSKRTYVLNESFDTC | 240 | |
| QY | 241 | TLPNGWTMDADGDGHNLSTINVTATHTGDGAFSKSWTASGGAIDLSPDNYLVT | 300 | |
| DB | 241 | TLPNGWTMDADGDGHNLSTINVTATHTGDGAFSKSWTASGGAIDLSPDNYLVT | 300 | |
| QY | 301 | KVTVPNGKLSYVWSSQVPTNEHYGVFLSTTGNEANFTIKLLEETGLSDKPA | 360 | |
| DB | 301 | KVTVPNGKLSYVWSSQVPTNEHYGVFLSTTGNEANFTIKLLEETGLSDKPA | 360 | |
| QY | 361 | SEGKLPAPYQERTIDLSAYAGQVYLAHFHNSGTGIFRLYDLDVAVSGGSNDYTYV | 420 | |
| DB | 361 | SEGKLPAPYQERTIDLSAYAGQVYLAHFHNSGTGIFRLYDLDVAVSGGSNDYTYV | 420 | |
| QY | 421 | YRDNVIAQLAATTPNQENAVPGQVNYCEVKYTAGVSPKCKVTVVSGSNEFAHVN | 480 | |

Db 421 YRDNVNVIAQNLAAATTENQENAVPQYNYCUEVKYTAGVSPKVKCDVTVESNEFAHVQNL 480

QY 481 TGSVAGQKVTLLKWDAPN 497

Db 481 TGSVAGQKVTLLKWDAPN 497

RESULT 2

S49763

gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)

C:Species: Porphyromonas gingivalis

C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C:Accession: S49763

R:Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.

submitted to the EMBL Data Library, November 1994

A:Description: Cloning, sequence analysis and expression in *Escherichia coli* of prpR1 of

A:Reference number: S49763

A:Accession: S49763

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1526 <ADU>

A:Cross-references: EMBL:X82680

C:Genetics:

A:Gene: prpR1

C:Keywords: cysteine proteinase; hydrolase

Query Match 36.0%; Score 950.5; DB 2; Length 1526;

Best Local Similarity 31.8%; Pred. No. 1.1e-53;

Matches 236; Conservative 77; Mismatches 154; Indels 275; Gaps 20;

QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPAGMARILLEAH 68

Db 671 TATTGQKVTLLKWDAPSTKTATNTARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730

QY 69 DWEDGTGYQMLWDADHNNQYCASIPEE--SPWFANGTIPAGLYDPPEYKVPVNDASFP 126

Db 731 DWNDGSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

QY 127 TNFVLDTGASADIPAGTYDVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185

Db 790 TNWMDGTASVNIIPAGTYDFAIAAQAQNAKIWIAGGPTKEDDYFEAGKYHFLMKKMG 849

QY 186 PGDAASV-----

Db 850 SGDGTGLTISGGGSDYTYTVYRDGTIKIKEGLTATTFEDGVAGNHCYCEVKYTAGVS 909

QY 194 -----VTGEGGNEFAPVQNLQWSVGQTVTLTWOAPAS-----DKRTYVLNE 235

Db 910 PKVKCDVTVESNEFAPVQNLQWSVGQKVTLLKWDAPNGTTPNPNPNPNPGTTTISE 969

QY 236 SFDTQTLFNGWTMIDADGDGHNWLSITVINYNTATHTGDCAMPSKSWTASGGAKIDLSPDN 295

Db 970 SFE-NGIPASWKTIDADGDGHCW-KPGNAPGIAGYNSNGCVYSESF-GUGGIGV-LTPDN 1025

QY 236 YLVTPKVTPENGKLSYVSSQ--VPWTEHYGVFLSTGTGNEAANFTIKLLETGLS----- 350

Db 1026 YLITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASNFTALLEETITAKGVR 1085

QY 351 -----DKPA-----

Db 1086 SPEATRGRIQSTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTTTF 1145

QY 355 -----

Db 1146 ESSTHGEATAEWTTIDADGDGGLCLSSGQLDWTLAHGGINVSSFSWNGMALNPDNYL 1205

QY 360 -----

Db 1206 ISKDVGTATKVKYVAVNDGFFGDHYAVMISKTGTNAGDFTVWFETPENGKKGARRGL 1265

QY 360 --KSEGKVLPAPOBERTIDLSAYAGQVYLAFRHNSGTICFLYLDDV--AVSGGSSND 415

Db 1266 STEADCAKPSQWIBERTVLDLP--AGTK-YVAFRHYNSDLNLYILLDDDIQFTMGGSPTPTD 1322

QY 416 YTYTVYRNVNVIAQNLAAATTENQENAVPQYNYCUEVKYTAGVSPKVKCDVTVESNEFA 475

Db 1323 YTYTVYRDGTIKIKEGLTETTFEEDGVATGNHCYCEVKYTAGVSPKVKCVNVTV-NSTQPN 1381

QY 476 HVQNLTGSAVGQKVTLLKWDAPN 497

Db 1382 PVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 3

A55426

gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis

N:Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP;

C:Species: Porphyromonas gingivalis

C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999

C:Accession: A55426; D53113

R:Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, J. Biol. Chem. 270, 1007-1010, 1995

A:Title: Molecular cloning and structural characterization of the Arg-gingipain protein

A:Reference number: A55426; MUID:95138080; PMID:7836351

A:Accession: A55426

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1704 <PAV>

A:Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068

R:Pike, R.; McGraw, W.; Potempa, J.; Travis, J.

J. Biol. Chem. 269, 406-411, 1994

A:Title: Lysoine- and arginine-specific proteinases from Porphyromonas gingivalis. Isola

A:Reference number: A53113; MUID:94103245; PMID:8276827

A:Accession: D53113

A>Status: preliminary

A:Molecule type: protein

A:Residues: 228-249 <PIK>

A:Experimental source: H66

A>Note: sequence extracted from NCBI backbone (NCBIP:141694)

C:Keywords: cysteine proteinase; hydrolase

Query Match 35.8%; Score 944.5; DB 2; Length 1704;

Best Local Similarity 31.9%; Pred. No. 3.2e-53;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPAGMARILLEAH 68

Db 671 TATTGQKVTLLKWDAPSTKTATNTARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730

QY 69 DWEDGTGYQMLWDADHNNQYCASIPEE--SPWFANGTIPAGLYDPPEYKVPVNDASFP 126

Db 731 DWNDGSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPFEYTVPENADPSCSP 789

QY 127 TNFVLDTGASADIPAGTYDVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185

Db 790 TNWMDGTASVNIIPAGTYDFAIAAQAQNAKIWIAGGPTKEDDYVFEAGKYHFLMKKMG 849

QY 186 PGDAASV-----

Db 850 SGDGTGLTISGGGSDYTYTVYRDGTIKIKEGLTATTFEDGVATGNHCYCEVKYTAGVS 909

QY 194 -----VTGEGGNEFAPVQNLQWSVGQTVTLTWOAPAS-----DKRTYVLNE 237

Db 910 PKVKCDVTVESNEFAPVQNLQWSVGQKVTLLKWDAPNGTTPNPNPNPNPGTTTISE 969

QY 238 DTOTLPLNGWTMIDADGDGHNWLSITVINYNTATHTGDCAMPSKSWTASGGAKIDLSPDNYL 297

Db 970 E-NGIPASWKTIDADGDGHCW-KPGNAPGIAGYNSNGCVYSESF-GUGGIGV-LTPDN 1025

QY 298 VTPKVTVPENGKLSYVSSQ--VPWTEHYGVFLSTGTGNEAANFTIKLLETGLS----- 350

Db 1026 ITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASNFTALLEETITAKGVR 1085

QY 351 -----DKPAPMNLV-----KSEGVKL----- 366

Db 1086 EAIRGRIGQSTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTTTFES 1145

```

QY 367 ----PAPYQERTID-----LSAYAGQOV----- 385
Db 1146 STHGEAPAEWTTIDADGGQGWCLSSGQDLWTFHAGTGVVVASPSWNGMALNPONYLIS 1205
QY 386 ----- 385
Db 1206 KDVTGATKVKYYAVNDGPPGSHYAVMIKTGTNAGDFTVVFEEPTNGINKGARFGLST 1265
QY 386 -----
Db 1266 EANGAKPOSVMIERTVLDLPAGTKYVAFRHYNCSDLNILLDDIQFTMGGSPTPTDTYTV 1325
QY 421 YRDNVVIAQNLAATFENAPGOYNCVEVKYTAGVSPKCKDVTVEGSENEFAHVONL 480
Db 1326 YRDGKTKIEGLTETTFEEDGVATGNHCEYCVKVTAGVSPKCVNVTI-NFTQFNPVKNL 1384
QY 481 TGSVAGQKVTUKWDAPN 497
Db 1385 KAQPDGDDVVLKWEAPS 1401

RESULT 4
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyrinomonas gingivalis
N:Alternate names: lysine-specific cysteine proteinase 1, 60K
C:Species: Porphyrinomonas gingivalis
C>Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30836; T30526; A53113
R:Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulskie-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A:Title: Analysis of the prtp gene encoding porphypain, a cysteine proteinase of Porphyrinomonas
A:Reference number: Z20895; MUID:96213011; PMID:8631659
A:Accession: T30836
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1732 <BAR>
A:Cross-references: EMBL:U42210; NID:G1314325; PID:G1314326; PIDN:AAB06565.1
R:Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z20896
A:Accession: T30837
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A:Cross-references: EMBL:U75366; NID:G2182811; PID:G2182812; PIDN:AAB60809.1
R:Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A:Title: IS195, an insertion sequence-like element associated with protease genes in Porphyromonas
A:Reference number: Z20844; MUID:98298016; PMID:9632563
A:Accession: T30526
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A:Cross-references: EMBL:AF017059; NID:G2738802; PID:G2738803; PIDN:AAC26523.1
R:Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A:Title: Lysine- and arginine-specific proteinases from Porphyrinomonas gingivalis. Isolation
A:Reference number: A53113; MUID:94103245; PMID:8275627
A:Accession: A53113
A>Status: preliminary
A:Molecule type: protein
A:Residues: 229-249 <PIK>
A:Experimental source: H66
A>Note: sequence extracted from NCBI backbone (NCBI:141690)
C:Genetics:
A:Gene: prtp; prtk
C:Keywords: cysteine proteinase; hydrolase

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1.3e-44;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

```

```

QY 22 TAAAGGPKT-----APSVTHQAVOKGIRTSVKQLRD-----PIPMAGHILILEA 67
Db 693 TATTOGQKVTUKWEAPS-----AKAEGSREVKRIKIGDLFTVIEPANDVRANEAKVILAA 747
QY 68 HDVWEDGTGYQMLDADHNOYCASIPESFFANGTI PAGLYDP--FEYKVPVNADASFSP 126
Db 748 DNWGDNTGYQLLDADHNTFGSVIPATGLF--TGTASSNLYSANFEYLVPANADPVVTT 806
QY 127 TNFVLDTGASADIAGTYDYVIINPNP--GLIYIVGEG--VSKGNDVVRAGKTYHFTV 181
Db 807 QNIIVTGGGVVIEGGVYDYCIITNPEPASGRMWIAGDGGNQPARYDDFTFEAGKKYTFM 866
QY 182 QROGPDGAASV----- 193
Db 867 FRAGMGDTMEVEDDSSPASVYTYVYRDGTKIKEGLTATTFEEDGVAAGNHEYCVVEKYT 926
QY 194 -----VTGEGNEFAFPVQNLQNSVSGTQVTLTWQAP-----ASDKRTYVLN 234
Db 927 AGVSPKVKCKDVTVEGSENEFAFVQNLGTSSVGQVTKLWDA PNGTNPNNPNPNPGTTL 986
QY 235 ESFTDTLPGNWTMIDADGHNWMLSTINVTNTHTGDDGAMFSGKSWTASGAKIDLSPD 294
Db 987 ESFE--NGIPASWKTIDADGGHGM--KPGNAPGIAGYNNGCVYSESF--GLGGIGV--LTPD 1042
QY 295 NLYVTPKVTVPENGKLSYVWSSQ--VPWTNHYGVFLSTTGTGNEAANFTIKLLEETLGS--- 350
Db 1043 NYLITPALDLPGSKLTFWVCAQADANVASEHYAVASTGNDASNTWALLEETITAKGV 1102
QY 351 -----DKPAPMNLV-----KSEGVKLI----- 366
Db 1103 RSPKAIIRIGTQWTKQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTET 1162
QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
Db 1163 FESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWTFHAGTGVVVASPSWNGMALNPON 1222
QY 386 ----- 385
Db 1223 LISKDVTKATKVKYYAVNDGPPGSHYAVMIKTGTNAGDFTVVFEEPTNGINKGARFG 1282
QY 386 -----YIAFRHFNSTGIFRLYLDV--AVSGEGSSNDYT 417
Db 1283 LSTEANGAKPOSVMIERTVLDLPAGTKYVAFRHYNCSDLNILLDDIQFTMGGSPTPTDT 1342
QY 418 YTVYRDNVVIAQNLAATFENAPGOYNCVEVKYTAGVSPKCKDVTVEGSENEFAHV 477
Db 1343 YTVYRDGTKIEGLTETTFEEDGVATGNHCEYCVKVTAGVSPKCKVDVTV-NSTQFNPV 1401
QY 478 QNLTKGSAV--GQKVTUKWDAP 496
Db 1402 QNLTKGSAV--GQKVTUKWDAP 496
Db 1402 QNLTKGSAV--GQKVTUKWDAP 496

RESULT 5
140229
arginyl endopeptidase - Porphyrinomonas gingivalis
C:Species: Porphyrinomonas gingivalis
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40229
R:Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A:Title: Structural characterization of arginylgipain, a novel arginine-specific cysteine
A:Reference number: I40229; MUID:95168884; PMID:7864651
A:Accession: I40229
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-991 <RES>
A:Cross-references: GB:D26470; NID:G927644; PIDN:BAA05484.1; PID:G927645

Query Match 18.7%; Score 494.5; DB 2; Length 991;
Best Local Similarity 25.5%; Pred. No. 2.9e-24;
Matches 124; Conservative 41; Mismatches 90; Indels 231; Gaps 7;

```

```
F:537,539/Binding site: substrate (Asp, Asn). #status predicted

Query Match          5.8%; Score 156; DB 1; Length 123;
Best Local Similarity 20.4%; Pred. No. 0.043;
Matches 92; Conservative 64; Mismatches 128; Indels 166; Gaps 23;

QY 106 AGLYDPEYKVPYNADASFPTNFVLDGTASADIPAGTYDYVIINPNPGIYYIVGEGVSK 165
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 22 ASLFSP-----PIRVFADDNINLVSG-----DFESGTIDGWIKQGNTPLAVTEQAIGQ 72

QY 166 GNDYVVWVWAKYTHFTVQRQP-----GDAAVS-----VVTGE-GGNEFPAPVNQLQ 209
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 73 YSMKVTVGRITGY-----EGPAYSLFKGMKGESYSVSLKVLRLVSGNSNPLITVTMFR 126

QY 210 WSVSGQTGV-PLTWQAPASD-----KRTYVLNESFD-----                238
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 127 EDDNGRHYDIIVNQKQVSEDSWTVSQTYTLDYIGTLKILYMVVEPDPFILEYIIDVVV 186

QY 239 -----TQTLFNGWMTMDADGHNWLSTIN-----VYNTATHGTGAMFSKS 280
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 187 TTONPIQVGNIANETFEENG-----SGMIGTGSSVVKAVYGVA-HSGDYSLLTTG 237

QY 281 WTAS-GGAKIDLSPDNVLTPKVTVPENGLSYVMWSQVPWNHEHGVFLSTTTCNEAANF 339
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 238 RTANWNGPSYDL-----TKKIVPGQQNVNDFWK-----FVNGNDTFQIKA 278

QY 340 TIKLLEETLGSDDKPAPM-----NLVKSEGVKLPAPYQERTIDLSAYAQVVYLAFPHFN 393
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 279 TVKRAT-----SDKDNYIQVNDFAVNKGWEITEKGSF---TLPVDYSGISIVVESQ--N 328

QY 394 STGIERLYLDDVAVSGESSNDYT-----YTVYRD-----                423
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 329 PT-LFYIDDFSVIGEISNNQITIQNDPLDYSVFKYDFFPIGAVDPSRLNDADPRAQL 386

QY 424 -----NWVIAQNLAATTFTQENVAPQYNY 448
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 387 TAKHFNNMLVAEN---AMKPESIQLPTEGNF 412


RESULT 7
F90696
hypothetical protein EC0542 [imported] - Escherichia coli (strain O157:H7, subspecies C); Species: Escherichia coli
C; Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C; Accession: F90696
R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Hirasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. Gaswara Res. 8, 11-29, 2001
A; Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 A; Reference number: A99629; MUID:21156231; PMID:11258796
A; Accession: F90696
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-5291 <RAY>
A; Cross-references: GB:BA000007; PIN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A; Experimental source: strain O157:H7, substrain RIMD 0509952
C; Genetics:
A; Gene: EC0542

Query Match          5.8%; Score 153; DB 2; Length 5291;
Best Local Similarity 24.3%; Pred. No. 0.52;
Matches 137; Conservative 70; Mismatches 202; Indels 154; Gaps 33;

QY 5 NSLIIFLAVLLSLLCCQTAAAQG--GPKTAPSVTHTQAVQKGIKTSKVKDLRDPFAGMA- 61
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 3303 NNGYILTATVSDLNGLNGSASKGVTDITAPVISFNTVAGDVVINNVHEHQIILSGTAT 3362

QY 62 -----RIILEAHDVWEDGTGYOMLWDADHNQYGASIIPEESF-WFANG--TIPAGLYDFF 112
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 3363 GAVAGDRLV-----VTIAQQQVVTSTDASN--WSVGVPASVSLGLADGTVTISATITD-- 3414

QY 113 EYKVPYNADASFPTNFVLDGTASADIPAGTYDYVIINPNPGIYYIVGEGVSKGNDYVVE 172
```

```

Db 3415 -----SAGNSTQTHNVQNTAAVSLSVST-----ISGDNLIN-----AAE 3450
QY 173 AGKTYHFTVORQGPDAASVWVT-----GGGNEEAPVQ-NLQMSYSGQTVTLTWQAPSD 227
Db 3451 AGSA--LTLISGTTGTFATGTVTVLLNGKYS--ATIQSNGSWVN---VPAADVAALSD 3503
QY 228 KRYVNLNESPFTQTLPGNWTMDADGDGHNWLSLTNV-YNTA-----THTGDGAMFSK 279
Db 3504 QTSYTVSASAQ-----DSAGNGSNSTQTHNVQNTAAVSLSVSTISGDNLIN-- 3551
QY 280 SWTAGGAKIDLK-----PDNYLVTPKTVVPENGK-LSYVSSQVPWT-----NE 323
Db 3552 ---AEGSALTLGSGTGFATGTVTVLLNGKYSATIQSNGSWVNPAADVAALSDGT 3608
QY 324 HYGVELS---TTGNEAANFTIKLEETILGSKDPAP-----MNLVKSQVGLKLPAPYQERTIDL 377
Db 3609 SYTVSASQDSAGNSAT-----ASRSVAVDLTAPVINSINTVSTD-RNLNAEQQPPUTL 3661
QY 378 ---SAYAGQV-----YIAFRHFNSTGIFRLYLDVAVSREGSSNDYTYTYVRDNV 425
Db 3662 NGSTSAEVGQTVTVTFGGKTYTATVAANGTWALNPVAVDLAALGQ----- 3707
QY 426 VIAQNLATTFNQENAVQAGQNYCYEVKVTAGVSKVCKDVTVEGSNEFAHVQNLTG--- 482
Db 3708 --AQITIASV-NDRAGNPGQATHALTD---TVAPTV-TIATVAGDDIINNAEQLAGQTI 3760
QY 483 -----SAVGQKVTL-----KWDA 495
Db 3761 SGTITAEVGVQTVTVFNQGTWSA 3783

RESULT 8
S72640
endo-1,4-beta-xylanase (BC 3.2.1.8) xynA precursor - Thermoanaerobacterium thermosulfur
C;Species: Thermoanaerobacterium thermosulfurigenes
C;Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 16-Jul-1999
C;Accession: S72640; S72622
R;Matuschek, M.; Sahm, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes E
A;Reference number: S72640
A;Accession: S72640
A;Molecule type: DNA
A;Residues: 1-1234 <MATU>
A;Cross-references: EMBL:U50952; NID:g1255235; PIDN:AA08046.1; PID:g1255238
A;Experimental source: strain EM1
R;Matuschek, M.; Sahm, K.; Zibat, A.; Bahl, H.
Mol. Gen. Genet. 252, 493-496, 1996
A;Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 tha
A;Reference number: S72621; MUID:97033555; PMID:8879252
A;Accession: S72622
A;Molecule type: DNA
A;Residues: 815-1234 <MAW>
A;Cross-references: EMBL:U50952
C;Genetics:
A;Gene: xynA
A;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C;Keywords: glycosidase; hydrolase
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-1234/Product: endo-1,4-beta-xylanase XynA #status predicted <MAT>
F;38-185/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1>
F;195-339/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
F;385-675/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F;682-853/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1>
F;857-1042/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
F;1055-1108/Domain: S-layer repeat homology <SLK>

Query Match 5.8%; Score 152; DB 2; Length 1234;
Best Local Similarity 20.0%; Pred. No. 0.078;
Matches 97; Conservative 66; Mismatches 138; Indels 184; Gaps 23;
QY 118 VNADASFPTNFVLDGTASADIPAGTYDYIINPNPGIIVGSGVSKGNDYVVEAGKTY 177

```

```

Db 28 ITPADADDITNKLVSNG---DFETGIDGWIKQGNPTLEVTEQALIGQSMKVTGRTQTY 83
QY 178 HFTVORQGP-----GDAASY-----VVTGE-GGNEFAPVQNLQMSVSGQTV-TILT 220
Db 84 -----EGPAYSFGLKQKMGESYNVSLKVLVSGQSSNPFFITVTMFREDDNGKHVDIV 137
QY 221 WQAPASDRKTYVVLNESFDTQTLPGNWTMDADGDGHNWLSLTNVNTATHTGDGAMFSKS 280
Db 138 WQKQVSE-----DSWTTVSGTYTLDTGT-----LKTLYMY----- 168
QY 281 WTASGAKIDLSPQNYLVTPKTVVP-----ENGLKSYVSSQVQVPWTEHYGV--- 327
Db 169 -VESPDPTLEYIVDDVTPQNPQVGNVINGFENGNTSGWGTGSSVVKAVGVVAHS 227
QY 328 ---FLSTTGNEA-----ANPTIKL-----EETLGSDKPAP----- 355
Db 228 GQYSLITGRTANWNGPSYDLTKIVPGQYQNVDFWKFVNGNDTEQIKATVKATSKDN 287
QY 356 -----MNLVKSQVGLKLPAPYQERTIDLAYAGQVYLAFRHFNSTGIFRLYLDVAVS 408
Db 288 YIQVNDFFVNNKGBWTEIKGSF---TLPVTDYSGVSIYVESQ---NPT--LBFYIDDFSVI 340
QY 409 GEGSSNDYT-----YTVYRD-----NVVIAQNLAA 433
Db 341 GEISNNQITIQNDIPDLYSVEKDYFFIGVAVDSSRLNDADPHAQILTAKHFNMLVAEN--- 397
QY 434 TTFNQENVAPQYNYCYE-----VKYTAGVSPKVKCDVTVEGSNEFAHVQNLTGSAVGQK 488
Db 398 -AMKPESLQPTGEGFTFDNADKIVDY-----EIAHNMKMRGH----- 433
QY 489 VTLKW 493
Db 434 -TLW 437

RESULT 9
C97012
probably cellulosomal scaffolding protein precursor, secreted, cellulose-binding and col
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97012
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1483 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78886.1; PID:g15023809; GSPDB:GN00169
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0910

Query Match 5.7%; Score 150; DB 2; Length 1483;
Best Local Similarity 20.2%; Pred. No. 0.14;
Matches 127; Conservative 81; Mismatches 213; Indels 208; Gaps 29;
QY 27 GGPRTAP-----SVTHQAVQKGIKTSKVQDLRDPPIAGMARITILEAHDVWEDGTGY- 77
Db 522 GTPVTPSQINVEGSSATDQPKIDLNGNLTAKDVQ-----SGKTLVQGTDTYVTDGIT 577
QY 78 -----QMLWDADHNOYGAS-----IPEESFWFANGTTPAGLYDPFEVKVPV 118
Db 578 LSQSYLAGLAGQVTLTLDNFNGGASQTITINVVKNETVLSVGTSGNPGD--TVKVPV 635
QY 119 N-----ADASFSTNPF-VLDGTASADIPAGT---YDYIINPNPGIIV--- 159
Db 636 TISQVSTFVGLICMDSYDASKFTVKDLVPLNTOLVKDTONYSFIVNTSTEGKISITFTDP 695
QY 160 -----GEGVSKGNDYVVEAGKT---YHFTVORQGPDAASVWVTGEGGNEFAPVQNL 208

```

```

Db      636  TLNYPISVDGLIAYLDLFIINSNATAGSALTV-----DPAITLIVADEND-----KDI 743
QY      209  QMSVSGQVILTWQAPASDKRTYVLNESFDT--QTLP-----NGWTMIDA-DGDGH 256
Db      744  KDAASNGKITVTGSAPV--VQSSVVNTSSVTDQNAPOQAVSITFNGTVDKVDKASG- 800
QY      257  NWLSTINVYNATHTFGDGAMFSKSWTAGGA-----KIDLSFDN-----YLVTPKVTPE 306
Db      801  ---NLTGKAGSYATSDGLTQSQSLATLAAGTYTYTIDFSAGNAGFTFVVVKGKTUVGS 857
QY      307  NGKLSYWSSS-----QVPWT-----NEHYGVFLSTTGNEAANFTIK-----LLBET- 347
Db      858  ATTLAVGTGVSAGKATVKEVPTISKVTPVGLICAEIDYDASKETVKDVLPTDVLKDDT 917
QY      348  -----LGSDDKAPMNL-----VKSGVKLPAPYQERTIDLSAYAGQQVYLAFRH 391
Db      918  NYSFIVNTSTPKISITFTDPLANYPISADGI--LAYLDFIINSNATAGDSALTV--- 971
QY      392  FNSTGIFRLYDDVAVSGEGSNDYT-----YVDRNVIAQNLAAFTFNQENVAPQYNYCEVKY-T 455
Db      972  NPSGFIADENDKDIQDAASNGKITVTGSTPVAENSVVNTSSVTDQNAPOQAVSITL 1030
QY      418  -----YTVDRNVIAQNLAAFTFNQENVAPQYNYCEVKY-T 455
Db      1031  NGNTITDVKDASGNTLKAGSYTFTTSDGILTSLYAT-----LAAGTYTYTVDPSAGN 1084
QY      456  AGVSPKVKCDVTVEGSEFAHVQNLTGSA 484
Db      1085  AGTIVVVKAKTVVSSATTLAVGTVSGKA 1113

RESULT 10
G86643
hypothetical protein yBef [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86643
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86643
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1983 <STO>
A:Cross-references: GB:AE005176; PID:g12723000; PID:AAK04249.1; GSPDB:GNO0146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yBef

Query Match          5.6%; Score 147; DB 2; Length 1983;
Best Local Similarity 21.7%; Pred.No. 0.32;
Matches 117; Conservative 77; Mismatches 181; Indels 164; Gaps 30;

QY      39  AVQKGIRTSVKQLRDPDPAGMARIILEAHDVWEDGTGYQMLWDADHNOYGASIPESFW 98
Db      628  ATQGI-----VKRIDGI-----DYNREIPWKIDINSG-----YW 660
QY      99  FANGTIPAGLYDPFEYKVPVNDASFSPTNF-VLDGTASADIPAGTYDYVVIINPNGLIY 157
Db      661  MENWSL-----EDKMSGELTFLNTFQIDKTAGNKVLSPT-EYTLIKTAGFSV 709
QY      158  IVGEGVSIGND--YVVEAGTYHTVQKQPG-----DAASVVVTGEGN-----200
Db      710  SFNSPLKEGTHHYKQIKYKTKFTSDTSDVINDSGHEGDIKFVNDASMTKDKNGGDHTNNDH 769
QY      201  -EPAPVQNLQWSVGGQVTLTWQAPASDKRTYVLNESFDTQTLPGNTWMTIDADGDGHNWL 259
Db      770  KEFKPIPPFY--NGQK-SGSYNA-TSKKITWTIAANFQOELSNA-SITDPISSDQNTV 824
QY      260  S-----TINVNT-----ATHTGGMFSKSWTAGSGAKIDLSPDN-----YLVTPKYT 303

```


QY 426 -----VIAQNLAAATTFNQENAVPGQVNY 448
 Db : : : : :
 594 AWTAVGFRKGASFTKEYAQYIAENRQVSYD-KTLKGTGPNF 636
 : : : : :
 RESULT 12
 A83412
 Hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: A83412
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, N.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; PMID:20437337; PMID:10984043
 A:Accession: A83412
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2468 <STO>
 A:Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AAG05263.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA1874

Query Match 5.2%; Score 138.5; DB 2; Length 2468;
 Best Local Similarity 23.7%; Pred. No. 1.6;
 Matches 107; Conservative 47; Mismatches 166; Indels 131; Gaps 23;

QY 103 TIPAGLYDPFEYKVPVNADASFPTNFVLDGTASADIPAGTYDYVVIINPNPGIIVIGEG 162
 Db : : : : :
 319 TDPAG-----NNSTPVTVEAP-----DTTA-----PAPATD-----VQVAPDG 351
 : : : : :
 QY 163 VS-KGNDYVVEAGKYHTFVQROGQDAASVVTGEGNEFAPVONLQWSVSGQVTLTW 221
 Db : : : : :
 352 SSVTGN---ABEGAT---VGVDTDGQDPTTVVVGPGGSFEVPLN--PPLTNGETVTVIV 404
 : : : : :
 QY 222 QAPASDKRTYVLYNLSFDQTLPN-----NTATH-TGDGAM-----FSKSWTASG 285
 Db : : : : :
 405 TDPAGNSSTPTAEAPDFDAPQVNASGSLVSTAGTAEAGTVITVDGNGNFIGQTSADAN 464
 : : : : :
 QY 257 -NWLSL-----INNY-----NTATH-TGDGAM-----FSKSWTASG 285
 Db : : : : :
 465 GNWSFTPGSQLPDGTVVNVVARDAGNSSPATSIIVDGVAPNAPVVEPSNGSELGTAEP 524
 : : : : :
 QY 286 GAKIDLSPDNLVTPKVTVPENGLSYWVSSQVPTNEHYGVFLSTTGNENANFTIKLLE 345
 Db : : : : :
 525 GSSVLTLDGNGNPIGQTADANGWSFTPTPLP-----DGTVVNVVARDAGNSSPAS 579
 : : : : :
 QY 346 ETLGSDKPAFMLVKSEGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVV 405
 Db : : : : :
 580 VTVDVAVAPATPTVDPNSGT-----TLSTGTAEPGSSVTLTDGNGNPIG----- 621
 : : : : :
 QY 406 AVSCEGSSNDYTT-----VYRDNVIAQNLAAATTFNQENAVPGQVNYCVEKYTAGVSPK 461
 Db : : : : :
 622 QVTDAGSGN-WTFPTPLPNGTVV--NAPATDPSGNASSPAS-----VTVDVAVAPATPV 673
 : : : : :
 QY 462 V--CKDVTVEGSEFAHVQNLV-----GSAVQV 487
 Db : : : : :
 674 VNPSNGTTLSTGTAEPGATVTLTDGNGNPIGQ 704
 : : : : :
 RESULT 13
 S19011
 endo-1,4-beta-xylanase (EC 3.2.1.8) - Bacillus polymyxa
 C:Species: Bacillus polymyxa
 C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 15-Oct-1999
 C:Accession: S19011
 R:Gosalbes, M.J.; Perez-Gonzalez, J.A.; Gonzalez, R.; Navarro, A.
 J. Bacteriol. 173, 7705-7710, 1991
 A:Title: Two beta-glycanase genes are clustered in Bacillus polymyxa: molecular cloning,

A:Reference number: S19011; MUID:92041687; PMID:1938968
 A:Accession: S19011
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-635 <GOS>
 A:Cross-references: EMBL:X57094; NID:948815; PIDN:CAA40378.1; PID:948816
 A>Note: the authors translated the codon GAA for residue 78 as Gly, CCT for residue 272
 C:Function:
 A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
 A:Pathway: xylan degradation
 C:Superfamily: Clostridium xylanase A repeat homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:408-502/Domain: Clostridium xylanase A repeat homology <CKA>

Query Match 5.2%; Score 137.5; DB 2; Length 635;
 Best Local Similarity 21.2%; Pred. No. 0.27;
 Matches 123; Conservative 65; Mismatches 176; Indels 217; Gaps 32;

QY 22 TAAAGGPKTAP---SVTHQAVQGIKTSKVKDLRDPAGMARILLAEHVDWBDGTGY 77
 Db : : : : :
 156 TADTPIGPWTPLGKALVTHST-----FGMAGVTWLFDPVAVLVDGDTGY 200
 : : : : :
 QY 78 QMLWDADHNOYGASIPESPFWANGTIPAGLYDPEYKV-----PYNADA 122
 Db : : : : :
 201 --LVS-----GGGIPNES-----DFASIANPKTARVILKGDWTSVIGSATIDAPY 245
 : : : : :
 QY 123 SFSPT-----NFVLDGTASADIPAGTYDYVII-NP-----NPG 154
 Db : : : : :
 246 LFEDSGIHKYKYYCYCINFA--GTHPQVPAEIGYMYVDNPMGPPTTKGHFLKNPY 303
 : : : : :
 QY 155 ILYIVG-----EGVSKNDYVVEAGKY---HFT-VORQPGDAA 190
 Db : : : : :
 304 TFFGVGGNNHAFNFKNEWVYVYHAQTVSKAQ---IGAGKGYRSPHINKLVHKEDGSIS 360
 : : : : :
 QY 191 SW--VTGEGNEFAPVONLQWSVSGQVTLTWQAPASDKRTYVLYNESFDQTLPLNGWTM 248
 Db : : : : :
 361 EVQGNMTG-----IAQLSNMNPYTRVEAETIAWQAGVTTEPT----- 397
 : : : : :
 QY 249 IDAGDGHNMISTINVTATHTGEGAMFSKSWTASGAK-----IDLSP 293
 Db : : : : :
 398 -QASGGP-----ISNLNVTN--IHNGDVIAGKADFGSAGAKTFKANVATNVGNGIEVRLDS 451
 : : : : :
 QY 294 DNYLVTPKVTVPENGLKSYM--VSSQVPTNEHYGVFLSTTGNENAN-----FTIKL 343
 Db : : : : :
 452 ETGPIVSLKVPSTGGMQTVREVEITNNATGVNIYLVFTGSGGNNILNLDAMQFTDPT 511
 : : : : :
 QY 344 LEETLGSDDKPAFMLVKSEGVKLPAPYQERTIDLSAYAGQVYLAFRHF--NSTGIFRLY 401
 Db : : : : :
 512 GGNITTKVEAENMKIGTYAGKISAPFD---GVALYANAD-VYSYQYFANST----- 560
 : : : : :
 QY 402 LDDVAVSGESSNDYTYVYRDNVVIAQNLAAATTFNQENAVPGQVNYCVEKYTAGVSPK 461
 Db : : : : :
 561 -HNISVRG-ASSNAGTAKV---DLVIG-----GVTVGSENF-----TGKTP 597
 : : : : :
 QY 462 VCKDVTVEGSEFAHVQNLATGSVAGQVTL-----KWDA 495
 Db : : : : :
 598 V-----QTLNIIHATGQEIKALTSDDGTWDA 626
 : : : : :
 RESULT 14
 A11489
 probable peptidoglycan bound protein (LPXTG motif) lin0457 [imported] - Listeria innocua
 C:Species: Listeria innocua
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: A11489
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Meek, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: A11489
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2013 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95689.1; PID:g16412898; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin0457

Query Match 5.2%; Score 137; DB 2; Length 2013;
Best Local Similarity 20.9%; Pred. No. 1.5;
Matches 103; Conservative 82; Mismatches 163; Indels 144; Gaps 30;

Qy 74 GTGTCM-LWDADHNOYGASIEES---PWFANGITPAGLYPPFYKVPVADASFSTNPF 129
Db 1185 GDDEIALFDKQKEVQASIKTNKGEFSFD---VAIKNPADFKLVKATP---TRNF 1237
Qy 130 VLDGTASADIP---AGTYDYVIINPGIIYIVGEGVSKGNDYVVEAGK-TYHFTVORQG 185
Db 1238 VY---SAKNPLFNMTKEYTLNSVFGV---GGVAE--IYITETSKPTTKIILDKAV 1286
Qy 186 PGDAASVVVTGEGNEFAPVQNLQWSV-----SGQTV-----TLTW 221
Db 1287 TPNA---ITTESSDEAEVTN-EMTVVDSNGTVVYSGTGNTIRIPNDEGTYIAKNTATD 1341
Qy 222 QA--PASDKRTYVLNESFDOTLPNGWTMIDADGDGH-----NWLSTINVTATHTGCD 273
Db 1342 EAGNTASDEKTFDIDYTVPTLTVNQ-----DASAEVNSTEANIIMKPLNVAARDTHDGN 1396
Qy 274 ---GAMFSK-SW-----TASGAK-----IDLSPDNLVTPKVTVPENG 308
Db 1397 ITPVDYSKVKWDVLGYTPVTATDASGNKATQTNLRIVDVTSPTLIT-----NN 1449
Qy 309 KLSWVSSQVPTWNEHYGVFLSTTGN-----AANFTIKLEETPLG 349
Db 1450 PLTYSIENMRKLTEQELYKAAGLIGDNYDLAPGOSVQPNKQPMVFTSNFS--TIFSDIA 1507
Qy 350 SDKP-----APNLVYKSEGV-LPAPYQERTTDLAYACQVYLAPRHENSGIFRLYLD 404
Db 1508 SVKFGQYQVQNLADSSGNQALPQTITNVDTMGFTVKANVSV-HVNTTKTBAEFPQD 1566
Qy 405 VAVSGEGSSNDYTVYRDNVVIAQNLAAATFNQENV---APGQYNYCVKVKYTAGVSPK 461
Db 1567 ARLDVTDNNDDTT-----NLIITSNFA-----EKNVLNKGKYEVTISATDTKG--NQ 1612
Qy 462 VKQVTVGSGNE 473
Db 1613 TTKETVQVSKD 1624
RESULT 15
A36734
bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text change 16-Jun-2000
C:Accession: A36734; A35131; A35750; S08223; JN0335; I39849; B69596; JU0084
R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 5520-5521, 1990
A:Reference number: A36734; MUID:90368623; PMID:2118514
A:Status: preliminary
A:Contents: erratum
A:Accession: A36734
A:Molecule type: DNA
A:Residues: 1-1433 <SLO>
A:Cross-references: GB:M29035; NID:g143307; PIDN:AA62679.1; PID:g143308
R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 1470-1477, 1990
A:Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning
A:Reference number: A35131; MUID:90170864; PMID:2106512
A:Accession: A35131
A:Molecule type: DNA
A:Residues: 1-365, 'S', 367-682, 'EIMP', 893, 'Q', 895-896 <SL2>
A:Cross-references: GB:M29035
A:Note: the authors translated the codon GAA for residue 545 as Leu

R:Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J. Biol. Chem. 265, 6845-6850, 1990
A:Title: Cloning, genetic organization, and characterization of a structural gene encoded:
A:Reference number: A35750; MUID:90216713; PMID:2108961
A:Accession: A35750
A:Molecule type: DNA
A:Residues: 1-392, 'V', 394-828, 'NIRTRLYSLKFCRSRHKSV' <WUA>
A:Cross-references: GB:J05400; NID:g142607; PIDN:AA83362.1; PID:g142609
A:Note: this sequence has been corrected
A:Accession: B35750
A:Molecule type: DNA
A:Residues: 876-935, 'CG' <WU2>
A:Cross-references: GB:J05400; NID:g142607; PIDN:AA83363.1; PID:g1119197
A:Note: this sequence has been corrected
R:Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A:Title: Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus subtilis.
A:Reference number: S08223; MUID:90174995; PMID:2106671
A:Accession: S08223
A:Molecule type: DNA
A:Residues: 1410-1433 <MAS>
A:Cross-references: EMBL:X17344; NID:g40165; PIDN:CAA35224.1; PID:g809661
R:Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
A:Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric]
A:Reference number: JN0335; MUID:93005071; PMID:1368833
A:Accession: JN0335
A:Molecule type: protein
A:Residues: 195-218, 'A' <KAT>
A:Note: source of this material was Bacillus subtilis (natto)
R:Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4955-4864, 1988
A:Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli
A:Reference number: I39846; MUID:89008108; PMID:3139638
A:Accession: I39849
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-211 <RES>
A:Cross-references: GB:M22630; NID:g142938; PIDN:AA22458.1; PID:g551705
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte:
C.; Bron, S.; Broutillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch:
A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle:
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, F.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Seguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, ;
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: B69596
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1433 <KUN>
A:Cross-references: GB:Z99111; GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAB13404.1; P:
A:Experimental source: strain 168
C:Genetics:
A:Gene: bpr; bpf
A:Map position: 135 (degrees)
C:Superfamily: bacillopeptidase F; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-194/Domain: propeptide #status predicted <PRO>
F:195-1433/Product: bacillopeptidase F #status experimental <MAT>
F:218-466/Domain: subtilisin homology <SBI>
F:227,274,452/Active site: Asp, His, Ser #status predicted

Query Match 5.1%; Score 135.5; DB 1; Length 1433;

GenCore version 5.1.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:22:23 ; Search time 6.24316 Seconds
 (without alignments)
 4145.156 Million cell updates/sec

Title: US-08-353-485-2
 Perfect score: 2641
 Sequence: 1 MRKLSLFLSLVLLSLLCWG.....QNLTSAGVGKVTLKWDAPN 497

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 2641 | 100.0 | 2628 | 1 HGA2_PORGI | Q51845 porphyron |
| 2 | 2583 | 97.8 | 2164 | 1 HGA1_PORGI | P59915 porphyron |
| 3 | 494.5 | 18.7 | 991 | 1 CPGI_PORGI | P28784 porphyron |
| 4 | 472 | 17.9 | 989 | 1 PRTH_PORGI | P46071 porphyron |
| 5 | 156 | 5.9 | 1157 | 1 XNNA_THESA | P36917 thermoaer |
| 6 | 137.5 | 5.2 | 635 | 1 XIND_PABPO | P45796 paenibacill |
| 7 | 135.5 | 5.1 | 1433 | 1 SUBF_BACSU | P16397 bacillus su |
| 8 | 135 | 5.1 | 1655 | 1 OMPB_RICCN | Q9kka3 r outer mem |
| 9 | 131.5 | 5.0 | 725 | 1 YA33_SULSO | Q97297 sulfolobus |
| 10 | 131 | 5.0 | 1656 | 1 OMPB_RICJA | O66653 r outer mem |
| 11 | 129 | 4.9 | 794 | 1 OXAA_CHLGV | P59809 chlamydophi |
| 12 | 127 | 4.8 | 794 | 1 TOXA_CLODI | P16154 clostridium |
| 13 | 124 | 4.7 | 1616 | 1 SLAP_BACCI | P58284 bacillus ci |
| 14 | 121.5 | 4.6 | 972 | 1 CTAL_BACCI | P94286 bacillus ci |
| 15 | 120.5 | 4.6 | 587 | 1 P1CP_PESER | P42790 pseudomonas |
| 16 | 119 | 4.5 | 699 | 1 CH11_BACCI | P20533 bacillus ci |
| 17 | 118.5 | 4.5 | 1300 | 1 120K_RICRI | P14914 rickettsia |
| 18 | 118.5 | 4.5 | 1654 | 1 OMPB_RICRI | Q53047 r outer mem |
| 19 | 115.5 | 4.4 | 336 | 1 GPPI_SCHPO | P78958 schizosacch |
| 20 | 114.5 | 4.3 | 335 | 1 GPFC_GRAVE | P54270 gracilaria |
| 21 | 114.5 | 4.3 | 488 | 1 PHB_ALCFA | P12625 alcaligenes |
| 22 | 114.5 | 4.3 | 681 | 1 TKT2_YEAST | P33315 saccharomyc |
| 23 | 114.5 | 4.3 | 857 | 1 GELA_DICDI | P13466 dictyosteli |
| 24 | 114.5 | 4.3 | 2660 | 1 YBEJ_ECOS7 | Q8x8v7 escherichia |
| 25 | 113.5 | 4.3 | 1004 | 1 SUPO_BACBR | P09333 bacillus br |
| 26 | 113.5 | 4.3 | 1151 | 1 ITAI_HUMAN | P56199 homo sapien |
| 27 | 111 | 4.2 | 2201 | 1 TENA_HUMAN | P24821 homo sapien |
| 28 | 111 | 4.2 | 2358 | 1 YBEJ_ECOLI | P76347 escherichia |
| 29 | 110.5 | 4.2 | 408 | 1 G3PT_HUMAN | O14556 homo sapien |
| 30 | 110.5 | 4.2 | 2265 | 1 FINC_BOVIN | P07589 bos taurus |
| 31 | 110 | 4.2 | 642 | 1 FLID_CAMJE | Q9phw6 campylobact |
| 32 | 110 | 4.2 | 827 | 1 CSG_HALVO | P25062 halobacteri |
| 33 | 110 | 4.2 | 1409 | 1 HAP1_HAEIN | P44596 haemophilus |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 109.5 | 4.1 | 932 | 1 CDG3_HUMAN | Q9y3h0 homo sapien |
| 35 | 109.5 | 4.1 | 2021 | 1 OMPA_RICCN | Q52657 rickettsia |
| 36 | 109 | 4.1 | 591 | 1 FLGE_CAUCR | P35806 caulobacter |
| 37 | 109 | 4.1 | 917 | 1 SLAP_THETH | P35830 thermus the |
| 38 | 109 | 4.1 | 1120 | 1 STFR_ECOLI | P76072 escherichia |
| 39 | 108.5 | 4.1 | 337 | 1 G3P_PODAN | P32637 podospora a |
| 40 | 108 | 4.1 | 611 | 1 WDRI_CAEEL | Q11176 caenorhabdi |
| 41 | 108 | 4.1 | 828 | 1 MRKC_KLEPN | P21647 klebsiella |
| 42 | 108 | 4.1 | 2481 | 1 FINC_XENLA | Q91740 xenopus lae |
| 43 | 107.5 | 4.1 | 1475 | 1 APU_THETY | P16950 t amylopull |
| 44 | 107.5 | 4.1 | 2477 | 1 FINC_RAT | P04937 rattus norv |
| 45 | 107 | 4.1 | 666 | 1 NEPI_THEVU | Q60053 thermoactin |

ALIGNMENTS

RESULT 1

HGA2_PORGI STANDARD; PRT; 2628 AA.

AC Q51845;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Hemagglutinin A precursor.

GN HAGA.

OS Porphyromonas gingivalis (Bacteroides gingivalis).

OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;

OC Porphyromonadaceae; Porphyromonas.

OX NCBI_TaxID=837;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=381;

RX MEDLINE=97047672; PubMed=8926061;

RA Han N., Whitlock J., Progluske-Fox A.;

RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 391 contains four large, contiguous, direct repeats.";

RL Infect. Immun. 64:4000-4007(1996).

CC -!- FUNCTION: Agglutinates erythrocytes.

CC -!- SIMILARITY: Belongs to peptidase family C25.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcements> or send an email to license@sib-sib.ch).

CC -----

CC EMBL; U41807; AAB17128.1; -.

DR PIR; T28651; T28651.

KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.

FT SIGNAL 1 24 POTENTIAL

FT CHAIN 25 2628 HEMAGGLUTININ A.

FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.

FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.

FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.

FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.

FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.

SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 100.0%; Score 2641; DB 1; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 1.1e-162;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRKLSLFLSLVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVXDLRPIIPAGM 60

Db 1 MRKLSLFLSLVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVXDLRPIIPAGM 60

Qy 61 ARITLEAHVDWEDGTGYQMLWDADHNOYGASIPRESFWFANGTIPAGLYDFEYKVPVNA 120

Db 61 ARITLEAHVDWEDGTGYQMLWDADHNOYGASIPRESFWFANGTIPAGLYDFEYKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVLIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVLIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGTTLTWOAPASDKRTTVVLNESFDTQ 240
 DB 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGTTLTWOAPASDKRTTVVLNESFDTQ 240
 QY 241 TLPNGWTMIDAGDGHNLSTINVTNTATHTGDMGAMFSKSWTASGAKIDLSPNYLVT 300
 DB 241 TLPNGWTMIDAGDGHNLSTINVTNTATHTGDMGAMFSKSWTASGAKIDLSPNYLVT 300
 QY 301 KVTVPENGLSYWSSQVPWNEHYGVFLSTTGNAAFTIKLLEETIGSKDPAPMNLVK 360
 DB 301 KVTVPENGLSYWSSQVPWNEHYGVFLSTTGNAAFTIKLLEETIGSKDPAPMNLVK 360
 QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYV 420
 DB 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYV 420
 QY 421 YRDNVVIAQNLAAATFNQENVAPGQYNCVEVKYTAGVSPKVKCDVTVEGSEFAHVQNL 480
 DB 421 YRDNVVIAQNLAAATFNQENVAPGQYNCVEVKYTAGVSPKVKCDVTVEGSEFAHVQNL 480
 QY 481 TGSVAVGQVTKLWDAPN 497
 DB 481 TGSVAVGQVTKLWDAPN 497

RESULT 2

HGAL_PORGI
 ID AC P59915; STANDARD; PRT; 2164 AA.
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hemagglutinin A precursor.
 GN HAGA OR PG1837.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1] _
 RP STRAIN=W83;
 RC MEDLINE=22829867; PubMed=12949112;
 RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
 RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
 RA Dewhirst F.E., Fraser C.M.;
 RA "Complete genome sequence of the oral pathogenic bacterium
 RT Porphyromonas gingivalis strain W83."
 RL J. Bacteriol. 185:5591-5601(2003).
 CC -!- FUNCTION: Agglutinates erythrocytes (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE017178; AAQ66831.1; ALT_INT.
 DR TIGR; PG1837; .
 DR Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2164 HEMAGGLUTININ A.
 FT DOMAIN 26 539 PEPTIDASE C25-LIKE 1.
 FT DOMAIN 540 991 PEPTIDASE C25-LIKE 2.

FT DOMAIN 992 1443 PEPTIDASE C25-LIKE 3.
 SQ SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;
 Query Match 97.8%; Score 2583; DB 1; Length 2164;
 Best Local Similarity 98.0%; Pred. No. 4.9e-159;
 Matches 487; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 1 MRKLSLFSIAVLISLLCWCQTAAAGGPKTAPSVTHQAVOKGIRTSKVXDLRDPPIAGM 60
 DB 1 MRKLSLFSIAVLISLLCWCQTAAAGGPKTAPSVTHQAVOKGIRTSKVXDLRDPPIAGM 60
 QY 61 ARIILEARDVWEDGTGYQMLWDADHNOYGASIPSESEFWFANGTIPAGLYDFEYKVPVNA 120
 DB 61 ARIILEARDVWEDGTGYQMLWDADHNOYGASIPSESEFWFANGTIPAGLYDFEYKVPVNA 120
 QY 121 DASFSPTNFVLDGTASADIPAGTYDYVLIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVLIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGTTLTWOAPASDKRTTVVLNESFDTQ 240
 DB 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGTTLTWOAPASDKRTTVVLNESFDTQ 240
 QY 241 TLPNGWTMIDAGDGHNLSTINVTNTATHTGDMGAMFSKSWTASGAKIDLSPNYLVT 300
 DB 241 TLPNGWTMIDAGDGHNLSTINVTNTATHTGDMGAMFSKSWTASGAKIDLSPNYLVT 300
 QY 301 KVTVPENGLSYWSSQVPWNEHYGVFLSTTGNAAFTIKLLEETIGSKDPAPMNLVK 360
 DB 301 KVTVPENGLSYWSSQVPWNEHYGVFLSTTGNAAFTIKLLEETIGSKDPAPMNLVK 360
 QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYV 420
 DB 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYV 420
 QY 421 YRDNVVIAQNLAAATFNQENVAPGQYNCVEVKYTAGVSPKVKCDVTVEGSEFAHVQNL 480
 DB 421 YRDNVVIAQNLAAATFNQENVAPGQYNCVEVKYTAGVSPKVKCDVTVEGSEFAHVQNL 480
 QY 481 TGSVAVGQVTKLWDAPN 497
 DB 481 TGSVAVGQVTKLWDAPN 497

RESULT 3

CPGI_PORGI
 ID AC P28784; Q45168; STANDARD; PRT; 991 AA.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
 DE (RGP-1).
 GN RGP1.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
 RC STRAIN=381;
 RX MEDLINE=95168884; PubMed=7864651;
 RA Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
 RA Ikehara Y.;
 RA "Structural characterization of argingipain, a novel
 RT arginine-specific cysteine proteinase as a major periodontal
 RT pathogenic factor from Porphyromonas gingivalis."
 RL Arch. Biochem. Biophys. 316:917-925(1995).
 RN [2]
 RP SEQUENCE OF 228-270.
 RC STRAIN=HG66;
 RX MEDLINE=92406812; PubMed=1527017;
 RA Chen Z., Potempa J., Polancowski A., Wikstrom M., Travis J.;

RT "Purification and characterization of a 50-kDa cysteine proteinase
 RL (gingipain) from Porphyromonas gingivalis.";
 CC J. Biol. Chem. 267:18896-18901(1992).
 CC -!- FUNCTION: Thiol protease which is believed to participate in
 CC intracellular degradation and turnover of proteins. Its
 CC proteolytic activity is a major factor in both periodontal tissue
 CC destruction and in bacterial host defense mechanisms. Activates
 CC complement C3 and C5.
 CC -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
 CC immunoglobulins, with a preference for Arg in P1, and hydrophobic
 CC residues in P2 and P3.
 CC -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
 CC and/or Mg(2+) for stabilization. It is stimulated by glycine-
 CC containing dipeptides. It is resistant to inhibition by proteinase
 CC inhibitors in human plasma.
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D26470; BAA05484.1; -
 CC PUB; I40229; I40229.
 CC HSP; P95493; 1CVR.
 CC MEROPS; C25.001; -
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR001769; Peptidase C25.
 CC InterPro; IPR005336; Peptidase C25_C.
 CC Pfam; PF01364; Peptidase C25; 1.
 CC Pfam; PF03785; Peptidase C25_C; 1.
 CC Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
 KW SIGNAL 1 24
 FT PROPEP 25 227
 FT CHAIN 228 991 GINGIPAIN R1.
 FT ACT_SITE 438 438 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 471 471 NUCLEOPHILE (BY SIMILARITY).
 FT CONFLICT 264 265 RT -> TK (IN REF. 2).
 FT SEQUENCE 991 AA; 108782 MW; 03EE3F43CEBE2544 CRC64;
 SQ
 Query Match 18.7%; Score 494.5; DB 1; Length 991;
 Best Local Similarity 25.5%; Pred. No. 2e-24; Indels 231; Gaps 7;
 Matches 124; Conservative 41; Mismatches 90;
 QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR---TSKYKDLRDPAGMARIIIEAH 68
 DB 671 TATTGGQKVTLLKWDAPSTKTNTATARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730
 QY 69 DVWEDGTGYQMLWDADHNOVCASIPER--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
 DB 731 DVWMDGGYQILLDADHDQGVQVIPSDDTHLW--FNCSPVPAFLFAPFYTFEENADPSCSP 789
 QY 127 TNFVLDGTASADIAGYVDYVVIINPNPGI--TYIVGEGVSGKNDVYVAGKYHTTVORQG 185
 DB 790 TNIMMDGTASVNIAGYDYFAIAAPQANAKIWIAGQPTKEDDVFVAGKYHFLMKMG 849
 QY 186 PGDAASVVVTEGGEAFAPVQNLQWSVSGQVTTLTWQAPASDKRTYVLNESFDTQTLPNG 245
 DB 850 SGDCGTELTIS-----
 QY 246 WTMDADGDGHNLSTINVNTATHTGDMGAFSKSWTASGAKIDLSPDNVLTVPKVTVP 305
 DB 860 -----
 QY 306 ENGLSYWVSQVPTWNEHYGVFLSTTGNEAANFTKILEETGLGDKPAPMNLVKGSEVK 365
 DB 860 -----
 QY 366 LPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDLDDVAVSGEGSSNDYTYTVYRDV 425
 DB ||| : ||| |||

Db 860 -----EGGSDTYTYTYRDGT 875
 QY 426 VIAQLAATTFNOENVAPGQYNYCVKVTAGVSPKVKCKVTVBGSNEFAHVQNLTSAY 485
 Db 876 KIKEGLTETTYRDAGMSAQSHYCVKVAAGVSPKVCVDYIPDGVDVTAQKPVTLTV 935
 QY 486 GQKVTLL 491
 Db 936 GKTITV 941
 RESULT 4
 PRTH PORGI
 ID PRTH PORGI STANDARD; PRT; 989 AA.
 AC P46071;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 11-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protease prth (EC 3.4.22.-).
 GN PRTH.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=95012612; PubMed=7927685;
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RT "Cloning and characterization of a new protease gene (prth) from
 RL Porphyromonas gingivalis.";
 RL Infect. Immun. 62:4279-4286(1994).
 RN [2]
 RP ERRATUM.
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RL Infect. Immun. 62:5707-5707(1994).
 CC -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
 CC P. GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
 CC IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
 CC AND IS A VIRULENCE FACTOR.
 CC -!- SUBCELLULAR LOCATION: In membrane vesicles.
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L27483; AAA51298.1; -
 CC HSP; P23882; IFMT.
 CC MEROPS; C25.001; -
 CC InterPro; IPR002376; formyl transf.
 CC Pfam; PF00551; formyl transf. 1.
 CC Hydrolase; Thiol protease; Repeat; Virulence.
 FT REPEAT 270 323
 FT REPEAT 528 581
 SQ SEQUENCE 989 AA; 110238 MW; FA85FE8A3AC8944C CRC64;
 Query Match 17.9%; Score 472; DB 1; Length 989;
 Best Local Similarity 38.6%; Pred. No. 5.7e-23;
 Matches 119; Conservative 45; Mismatches 124; Indels 20; Gaps 10;
 QY 199 GNEFAPQNLQWSVSGQVTTLTWQAPASDKRTYVLNESFDTQT---LPNGWTMIDADGDG 255
 DB 58 GTEICCPSPLPKAPICSTSLMRLSKTNKADFTFTFESSTHGEAPAEWTTIDADGDG 117
 QY 256 HNW--LSTINVNTATHTGDMGAFSKSWTASGAKIDLSPDNVLTVPKVTVPENGKLSYW 313
 DB 118 QGMCLSSQLDMLTAHGGTNVVSFSWNG-----MALNPNDNLYLSKDVGTGATKVKYYA 172

GN XYND.
 OS Paenibacillus polymyxa (Bacillus polymyxa).
 OX Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
 NCBI_TaxID=1406;
 RN [1] _____
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 842;
 EX MEDLINE=92041687; PubMed=1938968;
 RA Gosalbes M.J., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
 RT "Two beta-glucanase genes are clustered in Bacillus polymyxa:
 RT molecular cloning, expression, and sequence analysis of genes
 RT encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
 RL J. Bacteriol. 173:7705-7710(1991).
 CC -!- FUNCTION: SHOWS XYLANASE ACTIVITY AS WELL AS ALPHA-L-
 CC ARABINOFURANOSIDASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
 CC linkages in xylans.
 CC -!- PATHWAY: Xylan degradation.
 CC -!- SIMILARITY: Belongs to family 43 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC _____
 DR EMBL; X57094; CAA04378.1; -.
 DR PIR; S19011; S19011.
 DR InterPro; IPR006594; CBD IV.
 DR InterPro; IPR005084; CBM_6.
 DR InterPro; IPR008979; Gal_bind_like.
 DR Pfam; PF03422; CBM_6; 1.
 DR SMART; SM00606; CBD IV; 1.
 KW Xylan degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL
 FT CHAIN 1 26
 FT CHAIN 27 635 ENDO-1,4-BETA-XYLANASE D.
 SQ SEQUENCE 635 AA; 67914 MW; F9DEC6996723316 CRC64;

 Query Match 5.2%; Score 137.5; DB 1; Length 635;
 Best Local Similarity 21.2%; Pred. No. 0.14;
 Matches 123; Conservative 65; Mismatches 176; Indels 217; Gaps 32;
 QY 22 TAAAGGKTAP-----SVTHQAVQKIRISKVLDLDPDPAGMARILLSEAHVDWEDGTGY 77
 DB 156 TADTPIGPWTDPGLKALVTHST-----PGNAGVTWLFDPFVILVDDGTGY 200
 QY 78 QMLWDADHNQYGASIPESFWFANGTIPAGLYDPFEYKV-----PVNADA 122
 DB 201 --LYS-----GGIPNES-----DPASIANPKTARVILKGDWTSVIGSATIIDAPY 245
 QY 123 SFSPST-----NFVLOGTAGADIPAGTYDYVII-NP-----NPG 154
 DB 246 LFEDSGIHKYNGKYYSYCINFA--GTHPQQYPAGEIGYVSDNPMGPFTYKGHFLKNPY 303
 QY 155 IYIVVG-----EGVSKGNDYVVEAGTKV---HFT-VQVQSGDAA 190
 DB 304 TPFVGGNHNAHVNFKNENYVYVYHACTVSKAO---IGAGKGYRSPHINKLVKEDGSIS 360
 QY 191 SVV--VTGEGNEFAPVQNLQWSVSGQTWLTWQAPASDKRTYVLNSESFTQPLPNGWTM 248
 DB 361 EVQGNMTG-----IAQLSNPNPYTRVEAEIAHQAGVTIEPT----- 397
 QY 249 IDAGDGHNLSTINVTNATHGTGDGMFSKSWTAGAK-----IDLSP 293
 DB 398 -QASGGP-----ISNLNVTN--IHNGDMVAVGKADFGSAGAKTFKANVATNMGNGNIEVLDS 451
 QY 294 DNVLVTPKVTVPENGLSYW--VSSQVPWTHNEHYVFLSTTNGEAN-----FTIKL 343
 DB 452 ETGFLVGLSKVPSTGGQMTVREVTETINNATGVHNIYLVFTSGSGLNLLMDAWQTPNT 511
 QY 344 LEETLGSDDKAPMNLVKSEGVKLPAPYQERTIDLSAYAGQOVYLAFRHF--NSTGIFRLY 401

Db 512 GGNITTKVEAENMKIGGYAGKISAPFD-----GVALYANAD-YVSYSQYFANST----- 560
 QY 402 LDDVAVSGEGSSNDYTYTYVDNVVIAONLAATTNQNENVAPGQYNYCVVEYKTAGVSPK 461
 DB 561 -HNISVRG-ASSNAGTAKV---DLVIG-----GVTVGSENF-----TGKTPT 597
 QY 462 VKQDVTVGSSNEFAHVQNLTGSAVGQKVTL-----KWDA 495
 DB 598 V-----QTLNTHITATGQDEIKLALTSDDGTWDA 626

 RESULT 7
 SUBF BACSU
 ID SUBF BACSU STANDARD; PRT; 1433 AA.
 AC P16397;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
 DE (90 kDa serine proteinase).
 DE BPR OR BPF OR BSU15300.
 OS Bacillus subtilis.
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1] _____
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
 RX MEDLINE=90170864; PubMed=2106512;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
 RA Pero J.;
 RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
 RT and cloning of the gene.";
 RL J. Bacteriol. 172:1470-1477(1990).
 RN [2] _____
 RP REVISIONS.
 RC STRAIN=168;
 RX MEDLINE=90368623; PubMed=2118514;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
 RA Pero J.;
 RL J. Bacteriol. 172:5520-5521(1990).
 RN [3] _____
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90216713; PubMed=2108961;
 RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
 RT "Cloning, genetic organization, and characterization of a structural
 RT gene encoding bacillopeptidase F from Bacillus subtilis.";
 RL J. Biol. Chem. 265:6845-6850(1990).
 RN [4] _____
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessières P., Solotkin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruchli C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz K., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivoita C., Roche E., Roche E., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

DR SMART; SM00060; FN3; 3.
 KW Hypothetical protein; Kelch repeat; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 28 POTENTIAL
 FT CHAIN 29 725 HYPOTHETICAL PROTEIN SSO1033.
 FT REPEAT 59 100 KELCH 1.
 FT REPEAT 101 145 KELCH 2.
 FT REPEAT 146 199 KELCH 3.
 FT REPEAT 201 248 KELCH 4.
 FT REPEAT 250 297 KELCH 5.
 FT REPEAT 299 342 KELCH 6.
 FT DOMAIN 323 398 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 412 490 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 585 652 FIBRONECTIN TYPE-III 3.
 SQ SEQUENCE 725 AA; 78465 MW; FE8F122D0D013BAB6 CRC64;

Query Match 5.0%; Score 131.5; DB 1; Length 725;
 Best Local Similarity 20.4%; Pred. No. 0.4;
 Matches 112; Conservative 61; Mismatches 134; Indels 241; Gaps 27;

QY 101 NGTIPAGLYDFEYKVPVNADAFSPTN-----FVLDGTASADI----- 139
 Db 154 NTSFAGLYFP-----PSNAIRLFPNNDSWRILGYMPVPTGGYVFGTSLIIVSGYI 208
 QY 140 --PAGYDYVILNPN-----PGLIIVYEGVSKGNDYVVEAG 174
 Db 209 GYSAYTNDILYSPQNNNTILNGVLPYHDSALAYRGLFTVG-----GYIYTAG 261
 QY 175 K-----TYHFTVOROG-----PGDAASVY-----VTGEGNEFAFPQNLQ----- 209
 Db 262 SGGVNNAILAYNGNLQRVGLVPVYSAGYVQGNMLYLAGGSLSDVSALQLITFN 321
 QY 210 -----WSVSGQTVTTLTQAPASDKRTYVYNESFDQTLPNGWTMIDADGDGHNWL 259
 Db 322 FPLPPKITYSGAGNESVTLLGW-----NPVRLSSGYEIIYWNMGFN-- 363
 QY 260 STINVTNTAHT-----GPGAMF----- 277
 Db 364 SSINVGNTVYTVGLKDGITYFEVLAYNISGYSPSSIIALTPASVPNPQLVSVKYG 423
 QY 278 ----SKSW----TAGS-----GAKIDLSPDNVLV-----TPKVT----- 303
 Db 424 NDNVTLNWLPTFSGGVLGILGYVIVKNSWSSHFVNSTSLTISLNTVYNVFIYA 483
 QY 304 VPENGKLSYVWSQVPTNEHYGVFLSTTQNE-AANFTIKLEETLGSDDKPAFMLVKSE 362
 Db 484 VNLGNSPLVLTVPITTKASVFATIKLGNILVNTTSP-----PANIT-- 529
 QY 363 GVKLPAFYQERTIDLSAVAGQVYLAER-----HNSGTGER-----LYL-- 402
 Db 530 -LELYNPNGLNISQIALKGNSSYL-FRVPQGNVTLVLIASNSAGSVYVQVYVYLP 587
 QY 403 -----DDVAVSGESSNDYTYTVYRDNVVIAQNLAAATTNQNENAVFGQYNYCDEV 452
 Db 588 SPOVSLIGFNGNLVYSNNEANVTILVYVNNSLVYEGSPNSIVT--NISNG--TYLVKV 643
 QY 453 KYTAGVSP 460
 Db 644 ---IGVNP 648

RESULT 10
 OMPB RICUA STANDARD; PRF; 1656 AA.
 ID OMPB RICUA
 AC O06653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB
 OS Rickettsia japonica.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
 japonica."
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
 CC similarity).
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB003681; BAA20138.1; -
 CC InterPro: IPR006315; Autotransport.
 CC InterPro: IPR005546; Autotransporter.
 CC Pfam: PF03797; Autotransporter; 1.
 CC TIGRFAMs: TIGR01414; autotrans_bar1; 2.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 51.0%; Score 131; DB 1; Length 1656;
 Best Local Similarity 21.3%; Pred. No. 1.3;
 Matches 99; Conservative 45; Mismatches 198; Indels 122; Gaps 18;

QY 100 ANGTIPAGLYDFEYKVPVNADAFSPTNPNV-----LDGTASADIAGT----- 143
 Db 439 ANGTLAS-----ASADANVAVTNNTITAERASGVGVQLSGTHTAELRLNAGSV 487
 QY 144 ---YDYVINPNPGLIIVYEGVSKGNDYVVEAGKTVHFTVQRQGP-----DAA 190
 Db 488 FKLADGTGVNGKVNQTVLVG-GVLAAGAITLDSAITGIDNGGGGGAALQSITLANDAT 546
 QY 191 SVVVTGEGNEFAPVQNLQWSVSGQVTLTWQ-----APASDKRTYVYNESFDQT 240
 Db 547 KTLTLGANNIISANGGTINFQANGGTTKLTSTQNINWVDCDLAIATDQTGV-----DAS 601
 QY 241 TLPNGWTM-----IDADGDGHNWLSTINVTATHTGDNAMFSKSWTASGAKIDLSPDN 295
 Db 602 SLTNAQTLTISGTIGTIGIANNNTLQGFNQFNSKSTTLNGNVAINELVINGNNSVQFAHNT 661
 QY 296 YLVTPKVTPENGKLSYVWSQVPTNEHYGVFLSTTQNEAANFTIKLEETLGSDDKPA 355
 Db 662 YLITRTNNAAGQKILF-----NPVNNNTLAAAGTNLGSAN---PLAINEFS-KGAR 712
 QY 356 MNLVKS--EGVKLPAPYQERTIDLSAYAGQQVYLAFRHFNSTGIFRLYLDVAVSGB--G 411
 Db 713 ADIVLVNVEGVNL---YATNITTTDANVGSFV-----FNAGG-----KNIVSGTVGG 756
 QY 412 SSNDYTYTVYRDNVVIAQNLAAATTNQNENAVQYNYCDEVKYTA----- 456
 Db 757 QQGNKENTVALDNGTIVKFLGNATNGNTTIAANSTLQISGNTYADTFIASADGTGIVEFV 816
 QY 457 -----GVSPKVKCKDVTVEGS-----NEFAHVQNLTGS 483

817 NTGPNVTLNKOAVPVALKQITVSGPGRVNVVNEIGNAGNYHGA 860

```
Db 186 PGDAASVVVTGEGGNEPAPVQNLQMSVSGQTVTLTQWAPAD--KRTYVLNSESFDQTLP 243
Db 278 PSEASPGPLSVGANN-QPVSD-----TVGGYPLLRGILSDAKRT---PSSYHALNTV 329
QY 244 NGWTMIDADGDGHNWLSITVNTATHTGDDGAM-----FSKS 280
Db 330 SGRELTVNSVAGYR-VSTFNSMTLESDGSIKTKYKLPQOQYAFVEVGVNRSDDL 388
QY 281 WTASGAKIDSPDNYLVT-----KVTVE-----NGKLSYVWSQ 317
Db 389 WITSGIPEVEIMSNAF--TPAIKYHVIKKNQGLDKVLPKAKDPLALRSVYVQWILN- 445
QY 318 VPTWNEHGVFLSTTGNENANETIKLLEBTGLSGDKPAMNLYKSEG-----VK 365
Db 446 ----SNGYFGIILSPLTDPAGYAAAYVP---GSSVPTRLSLSLSPKNOQYPAKYPGYETL 499
QY 366 LPAPYQERTIDLSAYAGQOQVYLAFR-----HFNSTGIFRLYLDVAVSG 409
Db 500 LFLPQKEGTHRLVYAGPLADPTLRVLDKAYTNSKGESQYLDCTIFRG 548

RESULT 12
TOXA_CLODI STANDARD; PRT; 2710 AA.
ID TOXA_CLODI
AC P16154;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90221894; PubMed=2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A.";
RL Nucleic Acids Res. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90229305; PubMed=2105276;
RA Dove C.H., Wang S.-Z., Price S.B., Phelps C.J., Lysterly D.M.,
RA Wilkins T.W., Johnson J.L.;
RT "Molecular characterization of the Clostridium difficile toxin A
  gene.";
RL Infect. Immun. 58:480-488(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Only after the enteral delivery of the enterotoxin A may
  the characteristic disease called pseudomembranous colitis be
  induced.
CC -1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
  REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
  DIFFERENT OLIGOPEPTIDES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; X51797; CAA36094.1; -.
CC DR EMBL; M30307; AAA23283.1; -.
CC DR
```

```
QY 80 LWDADHNOYGASIPESFW-FANGTIPAGLYDPEYK-----VPVNADASFSPTNFVLG 133
Db 167 VYNKDSAVYGTSL---VFMSSGNEFLPLGYNSKEERLESIDLPTKAAVFSKSNAG 223
QY 134 TASAD--IPAGTYDYVINPNPGIYTVGEGV-----SKNDYVY-BAGKTYHTVORQG 185
Db 224 ANGAQYFVLNEMQLIVSQESGI-----EGINLPFSSDNKSTVNEIG--FDRELKAAQ 277

Query March 4.9%; Score 129; DB 1; Length 794;
Best Local Similarity 21.0%; Pred. No. 0.65; Indels 106; Gaps 21;
Matches 86; Conservative 63; Mismatches 154;

QY 80 LWDADHNOYGASIPESFW-FANGTIPAGLYDPEYK-----VPVNADASFSPTNFVLG 133
Db 167 VYNKDSAVYGTSL---VFMSSGNEFLPLGYNSKEERLESIDLPTKAAVFSKSNAG 223
QY 134 TASAD--IPAGTYDYVINPNPGIYTVGEGV-----SKNDYVY-BAGKTYHTVORQG 185
Db 224 ANGAQYFVLNEMQLIVSQESGI-----EGINLPFSSDNKSTVNEIG--FDRELKAAQ 277
```

DR EMBL: X92982; CAA63564.1; --
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF01473; CW binding_1; 28.
DR Pfam; PF04488; Gly_transf_sug; 1.
KW Toxin; Enterotoxin.
SQ SEQUENCE 2710 AA; 308052 MW; 0A6E52CEB4C14421 CRC64;

Query Match 4.8%; Score 127; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 4.4;
Matches 114; Conservative 66; Mismatches 200; Indels 178; Gaps 31;

QY 72 EDGQYQMLDADHNOYGAS-----IPESFWE-ANGTIPAGLY---DPEYK 115
DB 2117 EAATGWQTI---DGKKYFNTNTAISTGTYINGKHFFYNTDGMIGVFKGPNGEYF 2173

QY 116 VPVNADAS-----ESPTNFVLDGTA---SADIPAGTYDYVI-----INPNPGI--- 155
DB 2174 APANTDANNIEGQAILYQNEFLTLNGKKYFGSDSKAVTGWRIINNNKKYFNPNNALAAI 2233

QY 156 -----IYIVGEGYKNDYVVEAGKTYHFTVQROGPDGAASVVVG-----EGNE 201
DB 2234 HLCITINDKYYFSDGILQ-NGYITIBERNFYDANNE-----SKMVTGVFKGPNGEY 2286

QY 202 FAPVQNLQWSVGOTVT-----LTWQAPASDKRTYVILNESFDQTLPNGWTMIDADGDH 256
DB 2287 FAPANTHNNIEGQAIYVQNKFLT---LNGKKYFPDND---KAVTGWQTI-----DGK 2334

QY 257 NWLSTINVYNTAT--HFGDGAMF-----SKSWTASGAKIDLSPDNLYVTPKVTVP 305
DB 2335 KYTFNLNTAEATGWTQIDGKKYFNLNTAEATGWTQIDGKKYFNTNTFIASGTYS 2393

QY 306 ENGLKSYWSSQVPTWNEHGVFLSTG-----NEAAN-----FTIKLE 345
DB 2394 INGHFYNTDGI-----MQIGVFKGPNGEYFAPANTDANNIEGQAILYQNKFLTNGK 2449

QY 346 ETLSGSDKAPMNLVKSEGVKLPAFYQERTIDLSAYAGQ-----V 385
DB 2450 YFEGSDSKAVTGLTIDGK---YYFNTAVTGWQTINGKKYFNTNTSIASGT 2505

QY 386 YLAFRH--FNSGTGIFRLYL-----DDVAVSGESSNDYTVYVRDNY-VI 427
DB 2506 IISGKHFFYNTDGTIMQIGVFKGPDGFEYFAPANTDANNIEGQAILYQNRFLYLDNIYYF 2565

QY 428 AQNLAAAT-----ENQENVAPQYNY-CVEVK--YTAGVSPKCDVTVVEGSENF 474
DB 2566 GNNKAAATGWITDGNRYFEPNTAMGANGYKIDNKNFYRNLGLOI-----GVFKGSNGF 2622

QY 475 AH-----VQNLGTSAV 485
DB 2623 EYFAPANTDANNIEGQAI 2640

RESULT 13
ID SLAP_BACCI STANDARD; PRT; 1616 AA.
AC P35824;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE S-layer related protein precursor.
GN BUTB.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-3312;
RX MEDLINE=94374689; PubMed=7522196;
RA Aubert-Pivert E., Davies J.;
RT "Biosynthesis of butirosin in Bacillus circulans NRRL B3312:
RT identification by sequence analysis and insertional mutagenesis of
RT the butB gene involved in antibiotic production."

RL Gene 147:1-11 (1994).
CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. MAY PLAY A ROLE
CC IN THE EXPORT OF BUTIROSIDIN FROM THE ORGANISM.
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L20421; AAA62588.1; --
DR PIR; T17884; T17884.
DR InterPro; IPR003343; Big 2.
DR InterPro; IPR001119; SLH-
DR Pfam; PF02368; Big 2; 1.
DR Pfam; PF00395; SLH; 1.
DR SMART; SM00635; Bid 2; 1.
KW Signal; Cell wall; S-layer.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1616 S-LAYER RELATED PROTEIN.
SQ SEQUENCE 1616 AA; 172874 MW; 1F03FF7A4F113AA7 CRC64;

Query Match 4.7%; Score 124; DB 1; Length 1616;
Best Local Similarity 20.1%; Pred. No. 3.5;
Matches 120; Conservative 78; Mismatches 214; Indels 184; Gaps 31;

QY 1 MRKLSLFLSLVLLSL--CWGTAAAGQPKTAPSVTHQAVQKIGRTSKVKDLRD---P 55
DB 5 LRKNGGMIALVLSLITPAWGKASQSLQSGGV---QPQAGVTDDVYLSRDGTFR 61

QY 56 IPAGMARIILEAHVWEDGTGYQMLDADHNOYGASI---PEESFWFANGTIPAGLYDPF 112
DB 62 LPVGHINNVDAF---PQNTYVALFTS---GAQVTSQETDKVFVKKTALIAVDKH 113

QY 113 EYKVPNADASFPTNFVLDGTASADIPAGTY-----DYV 147
DB 114 DQVIRVIGTAVPTGTSWTENQNLPIEGGYVLLANDSSWGTSTVRKPLFEHYKTGDTV 173

QY 148 -----IINPNPGIIVIGEGYKNDYVVEAGKTYHFTVQROGPDGAASVVV 194
DB 174 SLHKGQVHVAGDFLAPDFGLNLVTSAGTT-----VTSPPDFTVSGQVVRVGSQGLSLTV 228

QY 195 TG-----EGGNEFAPVQNLQ--WVSQGTQVLTW---QAPASDKRTYV 232
DB 229 NGTEAALKADGAFQSAVRLTAGTNAISVKLLDKGREIVSVSTVTVYNDAAQQA- 281

QY 233 LNESFDQTLPNGWTMIDADGCHNWLSTINVTATHTGDMGFSKSWTASGAKIDLS 292
DB 282 --DLIEVEAAPIDIT-ISIEGPAHA-IGYVD-QDIAGIDDTVALFTNDW----- 325

QY 293 PDNLYVTPKVTVPENGKLSYWSSQVPTWNEHYGVFLSTTGNEAANFTIKLEETLGSCK 352
DB 326 -----GPQITVQ-----FNVAQV-----DAGSKVTKVNPSPIDGKT 358

QY 353 PA---PMNL-VKSGEVLKPAFYQERTIDLSAYAGQV--YLA----- 388
DB 359 PAWGTGTDLEIPSGGYVLA--QD-----TSYAGKNIKKYLATYFKVGDAILKRXGFAV 411

QY 389 -FRHENSTG--IFRLYLDVAVSSEGS-SNDYTYTVR---DNVIAQLAATNTNQEN 440
DB 412 PVKDLMTGTPFIARVTLNDNAMYETKPSLESGTITNMDPPSKIALTVNGTPELPFGPDG 471

QY 441 VAPQYNYCEVYKTAGVSPKCDVTVVEGSENF--VQNLGTSAVGQVTKW 493
DB 472 KFTSYTLASGINVLDIV-----VTEKGEQDSKDLVYVSRPFGSTGKKVIL-W 519

RESULT 14

```

CTAL1_BACCI
ID CTAL1_BACCI STANDARD; PRT; 972 AA.
AC P94286;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cycloisomaltooligosaccharide glucanotransferase precursor (EC 2.4.1.1-)
DE (CITASE).
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T-3040;
RA Oguma T., Kurokawa T., Tobe K., Kobayashi M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBCC databases.
CC -!- FUNCTION: PRODUCES CYCLOISOMALTOOLIGOSACCHARIDE FROM DEXTRAN.
CC -!- SIMILARITY: Belongs to family 66 of glycosyl hydrolases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D61382; BAA09604.1; -.
DR InterPro; IPR005084; CBM 6.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF03422; CBM 6; 2.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 38
FT CHAIN 39 972
FT GLUCANOTRANSFERASE.
FT SEQUENCE 972 AA; 107431 MW; 2827BBA61782CC22 CRC64;
SQ
Query Match 4.6%; Score 121.5; DB 1; Length 972;
Best Local Similarity 20.8%; Pred. No. 2.6;
Matches 71; Conservative 51; Mismatches 111; Indels 109; Gaps 17;
QY 183 RQPGDAASVVVTGEGN-----EFAPVQNLQSVS-----QQ--TVTLTWOA 223
Db 53 RYNPGDAVSIRYQAKNGTSSNGSAALEIFHLENSVTTSQSLSLTNGSQTLLFTWTA 112
QY 224 PASDKRTYVLNFSPTQTLPGNWTMDAGD-----GHWLSTINVTYATHTGDGAMES 278
Db 113 PSTDFRGYFVR--IDAGTLGQGATAIDVSDFTKPYGYISEF-----ESGETALES 163
QY 279 KWTASGGAKIDLSPDNYLVTPKVTVPENGKLSYVSVSQVPWNEHYGVFLSTTGNAAAN 338
Db 164 K-----AKVQLAQDY-----HINAWQFYDMWRHD----- 189
QY 339 FTIKLEETLGSDKAPMNLVSKGVKLPAPYQERTIDL-----SAYAGQOYVLAFRHF 392
Db 190 ---KMKIKRTGSSIDSTWLDLENRE---ISWSTLQIQIDAVHDVNGKAMAYAMIVASRENY 243
QY 393 NSTGI---FRLYLDVAVS-----GEGSSNDYTYTVYDNNVIVIAONLAATTFNQENVAP 443
Db 244 SPLGISPTGIYEDSHTNQDFDVGDSG---TYLYMSD-----PQN-----P 283
QY 444 GQNYCQVEKYTAGVSPKVKCVTVGSGNEFAHQVNLGTSAV 485
Db 284 NWQNY-IHAEYDINSINTAGDGHVDQMGQRNVVDYNGNSI 324

```

RESULT 15

P1CP_PSES

ID P1CP_PSES

AC P42790;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

DE Pseudomonalisin precursor (EC 3.4.21.100) (Pepstatin-insensitive
DE carboxyl proteinase) (Pseudomonapepsin).
GN PCP.
OS Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=33067;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 216-224.
RX MEDLINE=95014496; PubMed=7929375;
RA Oda K., Takahashi T., Tokuda Y., Shibano Y., Takahashi S.;
RT "Cloning, nucleotide sequence, and expression of an isovaleryl
RT pepstatin-insensitive carboxyl proteinase gene from Pseudomonas sp.
RT 101."
RL J. Biol. Chem. 269:26518-26524 (1994).
RN [2]
RP SEQUENCE OF 216-585.
RX MEDLINE=96157710; PubMed=8576087;
RA Hayashi K., Izu H., Oda K., Fukuhara K.-I., Matsuo M., Takano R.,
RA Hara S.;
RT "The primary structure of pepstatin-insensitive carboxyl proteinase
RT produced by Pseudomonas sp. No. 101."
RL J. Biochem. 118:738-744 (1995).
RN [3]
RP MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
RP SITES.
RX MEDLINE=99419069; PubMed=10488127;
RA Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
RT "Identification of catalytic residues of pepstatin-insensitive
RT carboxyl proteinases from prokaryotes by site-directed mutagenesis."
RL J. Biol. Chem. 274:27815-27822 (1999).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 216-585.
RX MEDLINE=21614391; PubMed=11747435;
RA Wlodawer A., Li M., Gustchina A., Dauter Z., Uchida K., Oyama H.,
RA Goldfarb N.E., Dunn B.M., Oda K.;
RT "Inhibitor complexes of the Pseudomonas serine-carboxyl proteinase."
RL Biochemistry 40:15602-15611 (2001).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.0 ANGSTROMS) OF 216-587.
RX MEDLINE=21223584; PubMed=11323721;
RA Wlodawer A., Li M., Dauter Z., Gustchina A., Uchida K., Oyama H.,
RA Dunn B.M., Oda K.;
RT "Carboxyl proteinase from Pseudomonas defines a novel family of
RT subtilisin-like enzymes."
RL Nat. Struct. Biol. 8:442-446 (2001).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the B chain of insulin at 13-
CC Glu-[Ala-14, 15-Leu]-Tyr-16 and 25-Phe-[Tyr-26 and angiotensin
CC I at 4-Tyr-[Ile-5. A good synthetic substrate is Lys-Pro-Ile-Glu-
CC Phe-[Phe(NO2)]-Arg-Leu.
CC -!- COFACTOR: Binds 1 calcium ion per subunit.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- PTM: Autocatalytically processed.
CC -!- SIMILARITY: Belongs to peptidase family S53.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D37970; BAA07188.1; -.
DR PDB; 1GA1; 11-FEB-03.
DR PDB; 1GA4; 11-FEB-03.
DR PDB; 1GA6; 11-FEB-03.
DR PDB; 1KDV; 11-FEB-03.
DR PDB; 1KDY; 11-FEB-03.
DR PDB; 1KDX; 05-FEB-03.
DR PDB; 1KE1; 11-FEB-03.
DR PDB; 1KE2; 11-FEB-03.
DR MEROPS; S53.001; -.
DR Pfam; PF00082; Peptidase_S8; 1.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:28:29 ; Search time 19.8443 Seconds
(without alignments)
7902.136 Million cell updates/sec

Title: US-08-353-485-2

Perfect score: 2641

Sequence: 1 MRKLSLFLSVLLSLCWG.....QNLGSAVGQKVLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 951.5 | 36.0 | 1706 | 2 Q51839 | Q51839 porphyromon |
| 2 | 950.5 | 36.0 | 1706 | 2 Q51838 | Q51838 porphyromon |
| 3 | 944.5 | 35.8 | 1704 | 2 Q51816 | Q51816 porphyromon |
| 4 | 943.5 | 35.7 | 1687 | 2 Q9R9B7 | Q9R9B7 porphyromon |
| 5 | 824 | 31.2 | 1723 | 2 P72194 | P72194 porphyromon |
| 6 | 818 | 31.0 | 1723 | 2 P72197 | P72197 porphyromon |
| 7 | 816 | 30.9 | 1358 | 2 P96967 | P96967 porphyromon |
| 8 | 812.5 | 30.8 | 1732 | 2 Q51817 | Q51817 porphyromon |
| 9 | 807.5 | 30.6 | 1732 | 2 Q52050 | Q52050 porphyromon |
| 10 | 806.5 | 30.5 | 1732 | 2 Q07442 | Q07442 porphyromon |
| 11 | 805.5 | 30.5 | 1223 | 2 Q9ZNB5 | Q9ZNB5 porphyromon |
| 12 | 561.5 | 21.3 | 1097 | 2 P72196 | P72196 porphyromon |
| 13 | 379 | 14.4 | 312 | 2 Q9KIB3 | Q9KIB3 porphyromon |
| 14 | 337 | 12.8 | 925 | 2 Q9F4J0 | Q9F4J0 porphyromon |
| 15 | 213.5 | 8.1 | 293 | 2 Q9XB09 | Q9XB09 porphyromon |
| 16 | 156 | 5.9 | 1454 | 16 Q8ENZ2 | Q8ENZ2 oceanobacil |

| | | | | | |
|----|-------|-----|------|-----------|---------------------|
| 17 | 154 | 5.8 | 1155 | 17 Q8TPZ8 | Q8TPZ8 methanosarc |
| 18 | 153 | 5.8 | 5291 | 16 Q8X2T1 | Q8X2T1 escherichia |
| 19 | 152 | 5.8 | 1234 | 2 Q60046 | Q60046 thermoanaer |
| 20 | 150 | 5.7 | 1483 | 16 Q977Y4 | Q977Y4 clostridium |
| 21 | 147 | 5.6 | 1983 | 16 Q9CJ55 | Q9CJ55 lactococcus |
| 22 | 146 | 5.5 | 2951 | 17 Q8THC9 | Q8THC9 methanosarc |
| 23 | 144.5 | 5.5 | 781 | 16 Q9S2G1 | Q9S2G1 streptomyce |
| 24 | 139 | 5.3 | 5020 | 16 Q8E9W3 | Q8E9W3 shewanella |
| 25 | 138.5 | 5.2 | 2468 | 16 Q9I2M3 | Q9I2M3 pseudomonas |
| 26 | 137.5 | 5.2 | 8173 | 16 Q7UDU8 | Q7UDU8 rhodospirell |
| 27 | 137 | 5.2 | 2013 | 16 Q9ZEK2 | Q9ZEK2 listeria in |
| 28 | 135 | 5.1 | 1615 | 2 Q9F0P9 | Q9F0P9 rickettsia |
| 29 | 135 | 5.1 | 1654 | 2 Q93QW9 | Q93QW9 rickettsia |
| 30 | 135 | 5.1 | 3064 | 16 Q82XT8 | Q82XT8 nitrosomona |
| 31 | 133 | 5.0 | 897 | 16 Q8G7K7 | Q8G7K7 bifidobacte |
| 32 | 133 | 5.0 | 1734 | 17 Q8TI67 | Q8TI67 methanosarc |
| 33 | 132 | 5.0 | 698 | 16 Q82F44 | Q82F44 streptomyce |
| 34 | 131.5 | 5.0 | 706 | 16 Q93J50 | Q93J50 streptomyce |
| 35 | 131.5 | 5.0 | 768 | 16 Q88G76 | Q88G76 pseudomonas |
| 36 | 131.5 | 5.0 | 1379 | 17 Q9HL55 | Q9HL55 thermoplasma |
| 37 | 131.5 | 5.0 | 1618 | 2 Q9KKB4 | Q9KKB4 rickettsia |
| 38 | 131 | 5.0 | 1357 | 17 Q8THC8 | Q8THC8 methanosarc |
| 39 | 131 | 5.0 | 1618 | 2 Q9KKB1 | Q9KKB1 rickettsia |
| 40 | 131 | 5.0 | 2009 | 16 Q7UY44 | Q7UY44 rhodospirell |
| 41 | 131 | 5.0 | 2768 | 16 Q8E9G6 | Q8E9G6 shewanella |
| 42 | 130.5 | 4.9 | 524 | 16 Q7U0W6 | Q7U0W6 rhodospirell |
| 43 | 130.5 | 4.9 | 1616 | 2 Q9KKB0 | Q9KKB0 rickettsia |
| 44 | 130.5 | 4.9 | 1617 | 2 Q9KKB9 | Q9KKB9 rickettsia |
| 45 | 130.5 | 4.9 | 2217 | 17 Q8TP72 | Q8TP72 methanosarc |

ALIGNMENTS

RESULT 1

Q51839 PRELIMINARY; PRT; 1706 AA.
AC Q51839; Q51840;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arginine-specific thiol protease precursor.
PRTR.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=95160709; PubMed=7857299;
RA Kirsbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakecki N.,
RA Reynolds E.C.;
RT "Complete nucleotide sequence of a gene ptrr of Porphyromonas
gingivalis W50 encoding a 132 kDa protein that contains an arginine-
specific thiol endopeptidase domain and a haemagglutinin domain.";
RL Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=96311339; PubMed=8713096;
RA Slakeski N., Cleal S.M., Reynolds E.C.;
RT "Characterization of a Porphyromonas gingivalis gene ptrr that encodes
an arginine-specific thiol proteinase and multiple adhesins.";
RL Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX Reynolds E.;
RT Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;

Db 790 TNNIMDGTASVNIPTAGTYDFAIAAPQANAKIWIAGQPTKEDDYVFEAGKYHFLMKMG 849
 QY 186 PGDAASV----- 193
 Db 850 SGDGTSLTISEGGSDYTVTVYRDGPKIKEGLTATTFEEDGVAAGNHEVCVEKYTAGVS 909
 QY 194 -----VTCEGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----DKRTYVLNE 235
 Db 910 PKVKDVTVEGNEFAPVQNLGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTISE 969
 QY 236 SEDTQTLPLNGWTMIDAGDGHNLWLTSTINYNTATHTGDMGFSKWTAGGAKIDLSPPN 295
 Db 970 SPE-NGIPASMTIDAGDGHGW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPDN 1025
 QY 296 YLVTVPKVTVPENGLSYWVSSQ-VPTWNEHYGVFLSTTGNENAFIKLEETLGS--- 350
 Db 1026 YLITPALDLPNGKLTFFWCAQDANYASEHYAVASSTGNDASFTNALLEETITAKVR 1085
 QY 351 -----DKPA----- 354
 Db 1086 SPEAIRGRIOSTWRQKTVLDLPAGTKYVAFRHQSTDMFYIDLDEVEIKANGKRAFDTTF 1145
 QY 355 -----PMNLV 359
 Db 1146 ESSSTHGEATAETITDADGDGQWCLSSGQDLWLTAGHGTNNVSSFSWNGMALNPDNYL 1205
 QY 360 ----- 359
 Db 1206 ISKDVGTATKVKYYAVNDGFGDHYAVMISKTGTNAGDFTVVFEETPNNGKGARFGL 1265
 QY 360 --KSEGKLPAPYQERTIDLSAYAGQVVLAFPHNSTGIFRLYLDV--AVSGEGSSND 415
 Db 1266 STEADGAKQSVWIERTVLP--AGTK-YVAFRHYNSDLNYILLDDIQFTMGSPPTPD 1322
 QY 416 YTVTVTRDNNVIAQNLAAATNENQENAPQYNYCVVEKYTAGVSPKVKDVTVEGNEFA 475
 Db 1323 YTVTVYRDGPKIKEGLTETTFEEDGVTAGNHEVCVEKYTAGVSPKVKCNVTV-NSTQFN 1381
 QY 476 HVONLTGSAVGQKVTLKWDAPN 497
 Db 1382 PVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 3

Q51816
 ID Q51816 PRELIMINARY; PRT; 1704 AA.
 AC Q51816;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arg-gingipain-1 proteinase.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95138080; PubMed=7836351;
 RA Favloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
 RA Travis J., Barr P.J.;
 RT "Molecular cloning and structural characterization of the Arg-
 RT gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
 RT proteinase-adhesin polypeptide."
 RL J. Biol. Chem. 270:1007-1010(1995).
 DR EMBL; U15282; AAA69539.1; --
 DR PIR; A55426; A55426.
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001; --
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.

DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase C25_C.
 DR Pfam; PF01364; Peptidase C25; 1.
 DR Pfam; PF03785; Peptidase C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 FT CHAIN 228 719 MATURE 50-KDA CYSTEINE PROTEINASE
 FT GINGIPAIN.
 SQ SEQUENCE 1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;
 Query Match 35.8%; Score 944.5; DB 2; Length 1704;
 Best Local Similarity 31.9%; Pred. No. 7.4e-50;
 Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;
 QY 22 TAAAGQ-----GPKTAPSYTHQAVQK--GIR--TSKVKDLRDIPIAGMARILEAH 68
 Db 671 TATTGQKVTLKWDAPSTKTNATTARSVDGIRLVLSSVSDAPELLRSGQAEIVLEAH 730
 QY 69 DWEDGTGOMLWDADHNOYGASIPPE--SFWANGTIPAGLYDPEYKVPFNADASF 126
 Db 731 DYVNDGSGYQILLDADHDQYGVIPSDHTLW-PNCSVPANLFAPEYTVPENADPSCP 789
 QY 127 TNVFLDGTASADI PACTYDYVLIINPNPI-IYIVGEGVSKNDYVVEAGKTVHFTVORQ 185
 Db 790 TNNIMDGTASVNIPTAGTYDFAIAAPQANAKIWIAGQPTKEDDYVFEAGKYHFLMKMG 849
 QY 186 PGDAASV----- 193
 Db 850 SGDGTSLTISEGGSDYTVTVYRDGPKIKEGLTATTFEEDGVTAGNHEVCVEKYTAGVS 909
 QY 194 -----VTCEGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----DKRTYVLNE 237
 Db 910 PKVKDVTVEGNEFAPVQNLGSAVGQKVTLKWDAPNGTPNPNPNPNPGTTTISESF 969
 QY 238 DTOTLPNGTMDADGDGHNLWLTSTINYNTATHTGDMGFSKSWTASGAKIDLSPDNYL 297
 Db 970 E-NGIPASMTIDAGDGHGW-KPGNAPGIAGYNSGCVYSEF-CLGGIGV-LTPDN 1025
 QY 298 VTPKVTVPENGLSYWVSSQ-VPTWNEHYGVFLSTTGNENAFIKLEETLGS----- 350
 Db 1026 ITPALDLPNGKLTFFWCAQDANYASEHYAVASSTGNDASFTNALLEETITAKVRSP 1085
 QY 351 -----DKPAPMNLV-----KSEGKLV----- 366
 Db 1086 EAIRGRIQSTWRQKTVLDLPAGTKYVAFRHQSTDMFYIDLDEVEIKANGKRAFDTTF 1145
 QY 367 ---PAPYQERTID-----LSAVAGQV----- 385
 Db 1146 STHGEAPAEWTITDADGDGQWCLSSGQDLWLTAGHGTNNVASFWSNGMALNPDNYL 1205
 QY 386 ----- 385
 Db 1206 KDVGTATKVKYYAVNDGFGDHYAVMISKTGTNAGDFTVVFEETPNNGKGARFGLST 1265
 QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTV 420
 Db 1266 EANGAKPQSVWIERTVLDLPAGTKYVAFRHYNSDLNYILLDDIQFTMGSPPTPDYTV 1325
 QY 421 YRDNNVIAQNLAAATNENQENAPQYNYCVVEKYTAGVSPKVKDVTVEGSEFAHVKQL 480
 Db 1326 YRDGPKIKEGLTETTFEEDGVTAGNHEVCVEKYTAGVSPKVCNVTV-NPTQFPVKNL 1384
 QY 481 TGSAGVQKVTLKWDAPN 497
 Db 1385 KAQPDGDDVVLKWEAPS 1401
 PRELIMINARY; PRT; 1687 AA.
 RESULT 4
 Q9R9B7
 ID Q9R9B7

AC Q9R9B7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hemagglutinin/protease.
 GN HAGE.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RA Han N., Dong H., Progulste-Fox A.;
 RT "Cloning and characterization of hgeS from P. gingivalis 381."
 RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF026946; AAD01810.1; -;
 DR HSP; P95493; 1CVR.
 DR MEROPS; C25.001; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64;
 Query Match 35.7%; Score 943.5; DB 2; Length 1687;
 Best Local Similarity 31.8%; Pred. No. 8.4e-50;
 Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;
 QY 22 TAAAGQ-----GPKTASVTHQVQK--GIR---TSKVKLDRDIPAGMARILLEAH 68
 DB 652 TATTQCKVTLKWDAPSTKTNATNTARSVDGIRELVLLSVDAPELLRSGQAEIVLEAH 711
 QY 69 DVMEGDTGYQLMDADHNOYGASTPER--SFWFANGTIPAGLYDPPEYKVPVNADASFSP 126
 DB 712 DVNWDGSGYQLLDADHDQGVIPSTHTLW--PNCVSPANLPAPFYITVPENADPSCSP 770
 QY 127 TNFVLDGTASADIPAGTYDYVIINPNFGI-IYIVGEGVSKGNDYVVBAGKTYHFTVQROG 185
 DB 771 TNMIMDGTASVNPAGTYDFAIAAPQANAKIWIAGQGTREDDYVFEAGKKYHFLMKMG 830
 QY 186 PGDAASVV----- 193
 DB 831 SGDGTETLISEGGSDYTYTYRDRGTKIKGLTATTFEEDGATGNHEYCVVVKYTAGVS 890
 QY 194 -----VTGGGNEFAPVQNLQWSVGQTVTLTWQAPAS-----DKRTYVLNE 235
 DB 891 PKVKDVTVEGSENEFAPVQNLGTSAGVQKVLKWDAPNGTPNPNPNPNPCTITLSE 950
 QY 236 SFDQTLPNGWMTDADGDGHNWLTSTINVTNATHTGDGAMFWSKTASGAKIDLSPDN 295
 DB 951 SFE-NGIPASNKTIADGDGHW-KPGNAPGIAGYNSGNCVYSESP-GLGGIGV-LTPDN 1006
 QY 296 YLVTPKVTPVNGKLSYVWSQ--VPWNEHYGVFLSTTGNEAANFTIKLEETLGS----- 350
 DB 1007 YLITPALDLPNGGKLTFWCAQDANYASEHYVYASSTGNDASNFTNALLEETITAKGVR 1066
 QY 351 -----DKFAPMNLV-----KSEGVKL----- 366
 DB 1067 SPEAIRGRIQGTWQKTVLDLPAGTKYVAFRHFQSTDMFYDLDRVEIKANGKRAFDTETP 1126
 QY 367 -----PAPYQERTID-----LSAYAGQV----- 385

DB 1127 ESSTHCEAPAEWTTIDADGDGQGNLCLSSGQDLWLTAHGTTNVASFWSNMGALNPDNYL 1186
 QY 386 ----- 385
 DB 1187 ISKDVGTGATKVKYKYAVNDGFPQDHYAVNMISKTGTNAGDFTVVFEETPNKGKGARFGI 1246
 QY 386 -----YLAFRHNSGTGIFRLYLDVV--AVSGEGSSNDYTY 418
 DB 1247 STEANGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNILLDDIQFTWGGSGFTDYTY 1306
 QY 419 TVVRNWNVIAQNLAAATTNQENVAPGOYNYCVVEVKYTAGVSKVCKDVTVEGSNEFAHVQ 478
 DB 1307 TVVRDGTGKIKEGLTETTFEEDGATGNHEYCVVVKYTAGVSKVCNVVVI-NPTQNPVK 1365
 QY 479 NLTGSAVGOKVILKWDAPN 497
 DB 1366 NLKAQPDGDDVVLKWEAPS 1384
 RESULT 5
 ID P72194 PRELIMINARY; PRT; 1723 AA.
 AC P72194;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Lys-gingipain.
 GN KGP.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RX MEDLINE=97044756; PubMed=8889827;
 RA Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
 RT "Cloning and sequencing of the gene encoding a novel lysine-specific
 RT cysteine proteinase (lys-gingipain) in Porphyromonas gingivalis:
 RT structural relationship with the arginine-specific cysteine proteinase
 RT (Arg-gingipain).";
 RL J. Biochem. 120:398-406(1996).
 DR EMBL; DB3258; BAA11870.1; -;
 DR MEROPS; C25.002; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25; 1.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1723 AA; 187261 MW; 5628963D251493EB CRC64;
 Query Match 31.2%; Score 824; DB 2; Length 1723;
 Best Local Similarity 29.8%; Pred. No. 2.3e-42;
 Matches 220; Conservative 75; Mismatches 170; Indels 274; Gaps 21;
 QY 22 TAAAGQPKT---ABSVTQAVQKQIRISKVQLRD-----PIPAGMARILLEA 67
 DB 693 TATTQCKVTLKWDAPS-----AKKAEASREVKRGIGLFTVIEPANDVANEAKVLLAA 747
 QY 68 HDVWEDGTGYQLMDADHNOYGASIPESFWFANGTIPAGLYDP-PEYKVPVNADASFSP 126
 DB 748 DNVWGTNTGYQLLDADHDNFTFGSVIPATGPLF-TGTASNLSYANFEYLIPANADPVVIT 806
 QY 127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIVIVGEG---VSKGNDYVVEAGKTYHFTV 181

```

Db 807 QNIIIVTQGVVPGVYDICTNPEPASGKMWIAGDGNQPARVDYDFTFAGKKYTFM 866
QY 182 QROGPGDAASVV----- 193
Db 867 RRAGMGDGTMEVEDDSSPASVYTYVYRDGTKEGLTATTFEEDGVAAGNHEYCVVEKYT 926
QY 194 -----VTGEGNEFAPQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 235
Db 927 AGVSPKVKCVTVEGNEFAPQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 986
QY 236 SFDTQTLNPGWMTIDADGDGHNWLTSTINVTATHTGDMFSGKSWTASGAKIDLSPDN 295
Db 987 SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPDN 1042
QY 296 YLVPKVTVPENGKLSYVWSQ-VPTWNEHYGVFLSTTGNAAFTIKLLETLGS----- 350
Db 1043 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVASSTGNDASFTNALLEETITAKGVR 1102
QY 351 -----DKPAPMLV-----KSEGVKL----- 366
Db 1103 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDMFYIDLDEVEIKANGKRAFDTFTF 1162
QY 367 -----PAPYQERTID-----LSAYAGQGV----- 385
Db 1163 ESSTHGEAPAEWTTIDADGDGQDMLCLSSGQLDMLTAHGNTNVVASFWSWNGMALNPDNVL 1222
QY 386 ----- 385
Db 1223 ISKQVTGATKVKYVAVNDGPFPGDHYAVMISKTGTNAGDFTVVFETPNKNGGARFGL 1282
QY 386 -----YLAFRHNSGTGIFRLYLDDV--AVSGEGSSNDYTY 418
Db 1283 STEANGAKPQSVWIBERTVDLPAGTKYVAFRHYNCSDNLVILLDDIQTFTMGSPPTDITY 1342
QY 419 TVYRDNVVIAQNLAAATTNQNENAVPQYNYCVVEKYTAGVSPKVKCVTVEGNEFAHVQ 478
Db 1343 TVYRDGTIKKEGLTETFEEDGVAATGNEHYCVVEKYTAGVSPKVKCVNTI-NPTQFNPK 1401
QY 479 NLTGSAVGQKVTWKWDAPN 497
Db 1402 NLKAQPDGDDVVVKWEAPS 1420

```

RESULT 6

```

P72197
ID P72197 PRELIMINARY; PRT; 1723 AA.
AC P72197;
DT 01-FEB-1997 (TremBLrel. 02, Created)
DT 01-FEB-1997 (TremBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Lys-gingipain.
GN KGP
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]_
RP SEQUENCE FROM N.A.
RA Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT "Molecular cloning and characterization of Porphyromonas gingivalis
RT Lys-gingipain."
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U54691; AAA99810.1; -.
DR MEROPS; C25.002; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

```

```

DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1723 AA; 186831 MW; 4508A7B50197CEBD CRC64;

Query Match 31.0%; Score 818; DB 2; Length 1723;
Best Local Similarity 29.4%; Pred. No. 5.4e-42;
Matches 218; Conservative 78; Mismatches 166; Indels 280; Gaps 21;

QY 22 TAAAGGGGKPT-----APSVTHQAVQKGIKRTSKVKDLRD-----PIPGAMRIILEA 67
Db 693 TATTGGQKVTWKWDAPS-----AKKAGSREVKRIGDGLFVTIEPANDVRANEAKVLA 747
QY 68 HDVWEDGTGYQMLWDADHINQIGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASPS 126
Db 748 DNVMGDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIIPANADPVVTT 806
QY 127 TNFVLDTGTSADIPAGTYDYVIINPNP--GLIYIVGEG--VSKGNDVYVVEAGKTYHFTV 181
Db 807 QNIIIVTQGVVPGVYDICTNPEPASGKMWIAGDGNQPARVDYDFTFAGKKYTFM 866
QY 182 QROGPGDAASVV----- 193
Db 867 RRAGMGDGTMEVEDDSSPASVYTYVYRDGTKEGLTATTFEEDGVAAGNHEYCVVEKYT 926
QY 194 -----VTGEGNEFAPQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 235
Db 927 AGVSPKVKCVTVEGNEFAPQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 986
QY 236 SFDTQTLNPGWMTIDADGDGHNWLTSTINVTATHTGDMFSGKSWTASGAKIDLSPDN 295
Db 987 SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPDN 1042
QY 296 YLVPKVTVPENGKLSYVWSQ-VPTWNEHYGVFLSTTGNAAFTIKLLETLGS----- 350
Db 1043 YLITPALDLPNGGKLTFWVCAQDANYASEHYAVASSTGNDASFTNALLEETITAKGVR 1102
QY 351 -----DKPA----- 354
Db 1103 SPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDMFYIDLDEVEIKANGKRAFDTFTF 1162
QY 355 -----PMLV 359
Db 1163 ESSTHGEAPAEWTTIDADGDGQMLCLSSGQLDMLTAHGNTNVVASFWSWNGMALNPDNVL 1222
QY 360 ----- 359
Db 1223 ISKQVTGATKVKYVAVNDGPFPGDHYAVMISKTGTNAGDFTVVFETPNKNGGARFGL 1282
QY 360 --KSEGVKLPAPOERTIDLSAYAGQVYLAFRHNSGTGIFRLYLDDV--AVSGEGSSND 415
Db 1283 STEANGAKPQSVWIBERTVDLP--AGTK-YVAFRHYNCSDNLVILLDDIQTFTMGSPPTD 1339
QY 416 YTYTVYRDNVVIAQNLAAATTNQNENAVPQYNYCVVEKYTAGVSPKVKCVTVEGNEFA 475
Db 1340 YTYTVYRDGTIKKEGLTETFEEDGVAATGNEHYCVVEKYTAGVSPKVKCVNTI-NPTQFN 1398
QY 476 HVQNLTGSAVGQKVTWKWDAPN 497
Db 1399 PVKNLKAQPDGDDVVVKWEAPS 1420

RESULT 7
P96967
ID P96967 PRELIMINARY; PRT; 1358 AA.
AC P96967;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hemagglutinin.

```

```

GN HAGD.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA Han N., Lepine G., Whitlock J., Wojciechowski L., Progulski-Fox A.;
RT "cloning, sequencing and characterization of hagd, a member of the
RT Harep multigene family in Porphyromonas gingivalis.";
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U68468; AAB49691.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003310; P:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

Query Match 30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%; Pred. No. 5.1e-42;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKTSKYKDLRD-----PIPGMARILIEA 67
DB 328 TATTQGGKVTLKWEAPS-----AKKAESREVRKIGDGLFVTIEPANDVRANEAKVVLAA 382
QY 68 HDVWEDGTGYQLMDADHNOYASIPESFWFANGTIPAGLYDP-PEYKVPVNADASFSP 126
DB 383 DNVWGDNTGYQLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVT 441
QY 127 TNFVLDTGASADIPAGTYDYVIINPNP--GIYIVGEG--VSKGNDYVVEAGTKYHFTV 181
DB 442 QNIIVTQGQEVVPGVGYDYCIITNPEPASGKMWIAGDGNQARYDDFTFEAGKKYFTFM 501
QY 182 QROGPGDAASV----- 193
DB 502 RRAGMGDGTDMVEDDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEVCVEVKYT 561
QY 194 -----VTGEGNEFAPVQNLQWSVSGVTTLTWOAPAS-----DKRTVYVLE 235
DB 562 AGVSPKVKDVTVEGSENEFAPVQNLGSAVGQKVTLKWDAPNGTNPNNPFGTTLSE 621
QY 236 SPTQTLPNGWTMIDAGDGHNLWLTINVNTNTHGDCGAMFSGKSWTASGAKIDLSPDN 295
DB 622 SFP-NGIPASWKTIDADGDGHG-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPDN 677
QY 296 YLVTPTKVTVPENCKLSVWSSQ-VPTWNEHYGVFLSTGTNEAANFTIKLEETLGS----- 350
DB 678 YLITPALDLANGKLTFWCAQDANVASEHYAVYASSTGNDASNTNALLEETITAKGVR 737
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
DB 738 SPEAIRGIQGTWRQKTVDLIPAGTKYVAFRHFQSTDMFYIDLDELVEIKANGRADTFET 797
QY 367 -----PAPQERTID-----LSAYAGQV----- 385
DB 798 ESSTHGEAPAEWTIDADGDGDQDLCLSGQLDLTAHGHTNNTVAGSFWNGMALNPENYL 857
QY 386 ----- 385
DB 858 ISKDVGTGATKVKYYAVNDGFGPDHYAVMLSKITGNAGDTFTVVFEETPNKGGARFGL 917

QY 386 -----YLAFPHNSTGIFRLYLDLV-----AVSGGSSNDITY 418
DB 918 STEANGAKPQSVMIERTVLDIPAGTKYVAFRHYNCSDLNLYLLDDITQFTWGGGFTPTDITY 977
QY 419 TVYRDNVIAQNLAAATTNQNENAPGOYNYCYVEVKYTAGVSPKVKCDVTVVEGSENEFAHVQ 478
DB 978 TVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKVCNVVTI-NPTQFNPK 1036
QY 479 NLTGSAVGQKVTLKWDAPN 497
DB 1037 NLKAQPDGDDGVVLKWEAPS 1055

RESULT 8
Q51817 PRELIMINARY; PRT; 1732 AA.
AC Q51817;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Porphyropain.
GN PRTP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W12;
RX MEDLINE=96213011; PubMed=8631659;
RA Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA Progulski-Fox A., Lantz M.S.;
RT "Analysis of the prtP gene encoding porphyropain, a cysteine proteinase
RT of Porphyromonas gingivalis.";
RL J. Bacteriol. 178:2734-2741(1996).
DR EMBL; U42210; AAB06565.1; -.
DR PIR; T30836; T30836.
DR MEROPS; C25_002; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003310; P:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCA4 CRC64;

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1.2e-41;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKTSKYKDLRD-----PIPGMARILIEA 67
DB 693 TATTQGGKVTLKWEAPS-----AKKAESREVRKIGDGLFVTIEPANDVRANEAKVVLAA 747
QY 68 HDVWEDGTGYQLMDADHNOYASIPESFWFANGTIPAGLYDP-PEYKVPVNADASFSP 126
DB 748 DNVWGDNTGYQLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVT 806
QY 127 TNFVLDTGASADIPAGTYDYVIINPNP--GIYIVGEG--VSKGNDYVVEAGTKYHFTV 181
DB 807 QNIIVTQGQEVVPGVGYDYCIITNPEPASGKMWIAGDGNQARYDDFTFEAGKKYFTFM 866
QY 182 QROGPGDAASV----- 193
DB 867 RRAGMGDGTDMVEDDDSPASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEVCVEVKYT 926

```

QY 194 -----VTGEGNEPAPVONLQWSVSGQTVTLTWQAP-----ASDKRTYVLN 234
 Db 927 AGVSPKVKCVDTVEGNEPAPVONLQSSVGQKVLKWDAPNGTPNPNPNPTTLIS 386
 QY 235 ESFDTQTLNPGMTMDADGGHNLSTINVTATHTGDMFSGKSWTASGGAKIDLSPD 294
 Db 987 ESFE-NGIPASAKTIDADGGHGW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPD 1042
 QY 295 NYLVTPEKVPENGLKSYWVSSQ-VPWTHYGVFLSTTGNBAANFTIKLEETLGS--- 350
 Db 1043 NYLITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDSNFTNALLEETITAKGV 1102
 QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
 Db 1103 RSPKAIIRIGTQWROKTVLDPAGTKYVAFRHFQSTDMFVLDLDEVEIKANGRADFTET 1162
 QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
 Db 1163 FESSTHGEAPAEWTTIDADGGQWLCLSSGQDLWLTAHGGSNVVSSFSWNGMALNPDNY 1222
 QY 386 ----- 385
 Db 1223 LISKQVGTGATKVKYYAVNDGPGDHVAVMISKTGNNAGDFTVVFETPNKNGGARFG 1282
 QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSNDYT 417
 Db 1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVLDDIQTWGGSPPTDYT 1342
 QY 418 YTVYRDNVVIQAOLAAATTNENQVAVGQYNYCEVVKYTAGVSPKCVKDVTVESGNEFAHV 477
 Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVKDVTV-NSTQFNPV 1401
 QY 478 QNLTGSV--GQKVTILKWDAP 496
 Db 1402 QNLTAEQAPNSMDAILKWNAP 1422

RESULT 9

OS2050
 ID OS2050 PRELIMINARY; PRT; 1732 AA.
 AC OS2050; (1)
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Lysine specific cysteine protease.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83.
 RX MEDLINE=98298016; PubMed=9632563;
 RA Lewis J.P., Macrina F.L.;
 RT "IS195, an insertion sequence-like element associated with protease
 RT genes in Porphyromonas gingivalis.";
 RL Infect. Immun. 66:3035-3042(1998).
 DR EMBL: AF017059; AAC26523.1;
 DR MEROPS: C25.002;
 DR GO: GO:000524; F:ATP binding; IEA.
 DR GO: GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO: GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO: GO:0006310; P:DNA recombination; IEA.
 DR GO: GO:0006281; P:DNA repair; IEA.
 DR GO: GO:0006260; P:DNA replication; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000977; DNA ligase.
 DR InterPro: IPR001769; Peptidase C25.
 DR InterPro: IPR005536; Peptidase C25_C.
 DR Pfam: PF01364; Peptidase C25_1.
 DR Pfam: PF03785; Peptidase C25_C; 1.
 DR PROSITE: PS00697; DNA_LIGASE_A1; 1.

KW Protease.
 SQ SEQUENCE 1732 AA; 187931 MW; B2337463D5CB5EA5 CRC64;
 Query Match 30.6%; Score 807.5; DB 2; Length 1732;
 Best Local Similarity 30.0%; Pred. No. 2.4e-41;
 Matches 222; Conservative 69; Mismatches 173; Indels 277; Gaps 22;
 QY 22 TAAAGGPKT---APSVTHQAVQGIKRTSKVKDLRD-----PIPAGMARILLEA 67
 Db 693 TATTQCKVTLKWEAPS-----AKRAEGSRVKKRIGDGLFVTIEPANDVRANEAKVVLAA 747
 QY 68 HDVWEDGTQYQLMDADHNYGASIPESFWFANGTIPAGLYDP-PEYKVPVNVNADASFSP 126
 Db 748 DNVWGDNTGYQLLDADHNTFGSVPATGPLF-TGTASSNLYSANPEYLVPANADPVVIT 806
 QY 127 TNFVLDGTASADIPAGTYDYVIINPNP--GIYIVGEG---VSKGNDYVVEAGKTYHTFTV 181
 Db 807 QNIIVTGGQEVVTPGGVYDYCITNPEPASGKMWIAGDGNQFARYDDTFEAGKTYFTM 866
 QY 182 ORQGGGDAASV----- 193
 Db 867 RRAGMGDGTMEVEDDPSASYTYTVYRDGTKIKEGLTATTFEDGVAAAGNHEVCVEKVT 926
 QY 194 -----VTGEGNEPAPVONLQWSVSGQTVTLTWQAP-----ASDKRTYVLN 234
 Db 927 AGVSPKVKCVDTVEGNEPAPVONLQSSVGQKVLKWDAPNGTPNPNPNPTTLIS 986
 QY 235 ESFDTQTLNPGMTMDADGGHNLSTINVTATHTGDMFSGKSWTASGGAKIDLSPD 294
 Db 987 ESFE-NGIPASAKTIDADGGHGW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPD 1042
 QY 295 NYLVTPEKVPENGLKSYWVSSQ-VPWTHYGVFLSTTGNBAANFTIKLEETLGS--- 350
 Db 1043 NYLITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDSNFTNALLEETITAKGV 1102
 QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
 Db 1103 RSPKAIIRIGTQWROKTVLDPAGTKYVAFRHFQSTDMFVLDLDEVEIKANGRADFTET 1162
 QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
 Db 1163 FESSTHGEAPAEWTTIDADGGQWLCLSSGQDLWLTAHGGSNVVSSFSWNGMALNPDNY 1222
 QY 386 ----- 385
 Db 1223 LISKQVGTGATKVKYYAVNDGPGDHVAVMISKTGNNAGDFTVVFETPNKNGGARFG 1282
 QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSNDYT 417
 Db 1283 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNVLDDIQTWGGSPPTDYT 1342
 QY 418 YTVYRDNVVIQAOLAAATTNENQVAVGQYNYCEVVKYTAGVSPKCVKDVTVESGNEFAHV 477
 Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVKDVTV-NSTQFNPV 1401
 QY 478 QNLTGSV--GQKVTILKWDAP 496
 Db 1402 QNLTAEQAPNSMDAILKWNAP 1422
 RESULT 10
 ID 007442
 AC 007442 PRELIMINARY; PRT; 1732 AA.
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Lysine-specific cysteine proteinase.
 GN PRK.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;


```

Db 478 ESFE-NGIPASWKTIDADGDGHW-KPGNAPGLAGYNSGCVYSSEF-GLGGIGV-LTPD 533
QY 295 NYLVTPKVTVPENGKLSYVWSQ-VPWTNHYGVFLSTGNEAANFTIKLLEETLGS----- 350
Db 534 NYLITPALDPLNGGKLTFFWVCAQDANYASEHVAVASSTGNDASNFTNALLEETITAKGV 593
QY 351 -----DKPAPMNLV-----KSEGKVL----- 366
Db 594 RSPEALRGIRIQGTWRQKTVLDLPAGTKYVAFRHFQSDTDMFYIDLDEVEIKANGKRADFTET 653
QY 367 -----PAPYQERTID-----LSAVAGQOV----- 385
Db 654 FESSTHGEAPAEWTITDADGGQGLCLSSGOLDLTAHGHTNVVASFWSNGMALNPONY 713
QY 386 ----- 385
Db 714 LISKDVTGATKYKYVAVNDGPFGBHYAVMISKTGTNAGDFTVFEETPNKNGKGARPG 773
QY 386 -----YLAFRHFNSTGIPRLYLDV--AVSGEGSSNDYT 417
Db 774 LSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYILLDIIQFTMGSPPTDYT 833
QY 418 YTVYEDNVVIAQNLAAFTNQENAPGQYKVEVKYTAGVSPKVKCKDVTVGSGNEFAHV 477
Db 834 YTVYRDGKIKEGLETTETFEEDGATGNHCEYCEVKYTAGVSPKVCVNTI-NPTQFNPV 892
QY 478 QNLTKSAV--GOKVTLKWDAP 496
Db 893 QNLTAEQAPNSMDAILKWNAP 913

RESULT 12
P72196
ID P72196 PRELIMINARY; PRT; 1097 AA.
AC P72196;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TonB-linked adhesin precursor.
GN TLA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=97386416; PubMed=9244265;
RA Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA Curtis M.A.;
RT "The tia gene of Porphyromonas gingivalis W50: a homologue of the
RT arginine-specific protease precursor (PrpRI) which shares sequence
RT similarity to TonB-linked receptors."
RL J. Bacteriol. 179:4778-4788 (1997).
RC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Y07618; CA68897.1; -
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006310; F:DNA recombination; IEA.
DR GO; GO:0006281; F:DNA repair; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 53 POTENTIAL.
SQ SEQUENCE 1097 AA; 118731 MW; 73BBA337B421F8B9 CRC64;

```

```

Query Match 21.3%; Score 561.5; DB 2; Length 1097;
Best Local Similarity 29.8%; Pred. No. 2.4e-26;
Matches 158; Conservative 48; Mismatches 117; Indels 207; Gaps 17;

QY 165 KNDYVVEAGKT---YHFT-----VQOQFGDAASVVVVTGEGNEFAPVQNLQWSVSG 214
Db 267 KGSDDYVATGLRFGIDFTPEWSNLNYQNVLGDAIPV---GGSNEFAPVQNLGSSVG 322
QY 215 QTVTLTWQAP-----ASDKRYVVLNESPDQTOTLNGWTMIDADGDGHNLSTINY 265
Db 323 QKVTLKNDAPNGTTPNPNPNPNTGTTLSSEF-NGIPASWKTIDADGDGHW-KPGNAP 380
QY 266 NTATITGDMGAFPSKSWTASGGAKIDLSPDNYLVTPKVTPENGKLSYVWSQ-VPWTNHY 324
Db 381 GIAGYNSGCVYSSEF-GLGGIGV-LTPDNYLITPALDPLNGGKLTFFWVCAQDANYASEH 438
QY 325 YGVFLSTGNEAANFTIKLLEETLGS-----DKPAPMNLV----- 359
Db 439 YAVYASSTGNDASNFTNALLEETITAKGVRSFKAIRIGTWRQKTVLDLPAGTKYVAFR 498
QY 360 -----KSEGKVL-----PAPYQERTID----- 376
Db 499 HFQSDTDMFYIDLDEVEIKANGKRADFTETFSSTHGEAPAEWTITDADGGQGLCLSSG 558
QY 377 ----LSAVAGQOV----- 385
Db 559 QLDNLTAHGSSNVVSSFSWSNGMALNPNDYLSKDVTKYKVVAVYVNDGPFGBHYAVMI 618
QY 386 -----YLAFR 390
Db 619 SKTGTNAGDFTVFEETPNKNGKGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFR 678
QY 391 HFNSTGIPRLYLDV--AVSGEGSSNDYTYYVYRDNVVIQNLAAFTNQENAPGQYNY 448
Db 679 HYNCSLDNYILLDDIIQFTMGSPPTDYTYYVYRDTGKIKEGLETTETFEEDGATGNHCEY 738
QY 449 CVEVKYTAGVSPKVKCKDVTVGSGNEFAHVQNLTKSAV--GOKVTLKWDAP 496
Db 739 CVEVKYTAGVSPKVCVNTI-NPTQFNPVQNLTAEQAPNSMDAILKWNAP 787

RESULT 13
Q9KIB3
ID Q9KIB3 PRELIMINARY; PRT; 312 AA.
AC Q9KIB3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237555; AAF81413.1; -.
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match 14.4%; Score 379; DB 2; Length 312;
Best Local Similarity 37.2%; Pred. No. 8.6e-16;
Matches 94; Conservative 44; Mismatches 85; Indels 30; Gaps 9;

```


Search completed: May 18, 2004, 11:35:05
Job time : 26.8443 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:30:39 ; Search time 8.24989 Seconds
(without alignments)
3110.116 Million cell updates/sec

Title: US-08-353-485-2

Perfect score: 2641

Sequence: 1 MRKLSLFLAVLLSLILWG.....QNLGSAVGQKVTWKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2.6/prodata/2/iaa/5A_COMB.pep.*

2: /cgn2.6/prodata/2/iaa/5B_COMB.pep.*

3: /cgn2.6/prodata/2/iaa/6A_COMB.pep.*

4: /cgn2.6/prodata/2/iaa/6B_COMB.pep.*

5: /cgn2.6/prodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2.6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 2641 | 100.0 | 497 | 2 | US-08-570-311-2 |
| 2 | 2641 | 100.0 | 497 | 2 | US-08-353-485-2 |
| 3 | 2641 | 100.0 | 2628 | 2 | US-08-570-311-14 |
| 4 | 951.5 | 36.0 | 1706 | 4 | US-09-066-330-10 |
| 5 | 944.5 | 35.8 | 1477 | 4 | US-09-482-500A-1 |
| 6 | 944.5 | 35.8 | 1704 | 3 | US-08-336-308A-10 |
| 7 | 944.5 | 35.8 | 1704 | 3 | US-08-822-324-6 |
| 8 | 944.5 | 35.8 | 1704 | 3 | US-09-490-931-10 |
| 9 | 943.5 | 35.7 | 1687 | 2 | US-08-570-311-29 |
| 10 | 816 | 30.9 | 1358 | 2 | US-08-570-311-27 |
| 11 | 812.5 | 30.8 | 1732 | 2 | US-08-570-311-10 |
| 12 | 812.5 | 30.8 | 1732 | 2 | US-08-353-485-10 |
| 13 | 806.5 | 30.5 | 1087 | 2 | US-08-570-311-8 |
| 14 | 806.5 | 30.5 | 1087 | 2 | US-08-353-485-8 |
| 15 | 806.5 | 30.5 | 1732 | 4 | US-09-066-330-11 |
| 16 | 546.5 | 20.7 | 450 | 2 | US-08-570-311-16 |
| 17 | 546.5 | 20.7 | 456 | 2 | US-08-570-311-18 |
| 18 | 546.5 | 20.7 | 456 | 2 | US-08-570-311-20 |
| 19 | 458.5 | 17.4 | 439 | 2 | US-08-570-311-22 |
| 20 | 405 | 15.3 | 942 | 1 | US-08-141-324-14 |
| 21 | 405 | 15.3 | 942 | 1 | US-08-541-902-14 |
| 22 | 175 | 6.6 | 49 | 3 | US-08-822-324-18 |
| 23 | 148 | 5.6 | 42 | 4 | US-09-066-330-3 |
| 24 | 140 | 5.3 | 2736 | 4 | US-09-252-991A-30227 |
| 25 | 135.5 | 5.1 | 2215 | 4 | US-09-543-681A-5434 |
| 26 | 132 | 5.0 | 509 | 3 | US-08-822-324-8 |
| 27 | 127 | 4.8 | 46 | 3 | US-08-822-324-9 |

| | | | | | | |
|----|-----|-----|------|---|-------------------|--------------------|
| 28 | 127 | 4.8 | 811 | 1 | US-08-480-604A-7 | Sequence 7, Appli |
| 29 | 127 | 4.8 | 811 | 2 | US-08-405-496A-7 | Sequence 7, Appli |
| 30 | 127 | 4.8 | 811 | 3 | US-08-915-136-7 | Sequence 7, Appli |
| 31 | 127 | 4.8 | 811 | 4 | US-08-957-310-7 | Sequence 7, Appli |
| 32 | 127 | 4.8 | 811 | 4 | US-10-011-366-7 | Sequence 7, Appli |
| 33 | 127 | 4.8 | 811 | 4 | US-09-084-517-7 | Sequence 7, Appli |
| 34 | 127 | 4.8 | 812 | 1 | US-08-480-604A-29 | Sequence 29, Appli |
| 35 | 127 | 4.8 | 812 | 3 | US-08-915-136-29 | Sequence 29, Appli |
| 36 | 127 | 4.8 | 812 | 4 | US-09-084-517-29 | Sequence 29, Appli |
| 37 | 127 | 4.8 | 2710 | 1 | US-08-480-604A-6 | Sequence 6, Appli |
| 38 | 127 | 4.8 | 2710 | 2 | US-08-405-496A-6 | Sequence 6, Appli |
| 39 | 127 | 4.8 | 2710 | 3 | US-08-915-136-6 | Sequence 6, Appli |
| 40 | 127 | 4.8 | 2710 | 4 | US-08-957-310-6 | Sequence 6, Appli |
| 41 | 127 | 4.8 | 2710 | 4 | US-10-011-366-6 | Sequence 6, Appli |
| 42 | 127 | 4.8 | 2710 | 4 | US-09-084-517-6 | Sequence 6, Appli |
| 43 | 126 | 4.8 | 1752 | 4 | US-09-865-621A-2 | Sequence 2, Appli |
| 44 | 123 | 4.7 | 951 | 4 | US-09-924-097A-15 | Sequence 15, Appli |
| 45 | 121 | 4.6 | 25 | 2 | US-08-902-516-46 | Sequence 46, Appli |

ALIGNMENTS

RESULT 1

US-08-570-311-2

; Sequence 2, Application US/08570311

; Patent No. 5824791

; GENERAL INFORMATION:

; APPLICANT: Progulske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guylaine

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ted W. Whitlock

; STREET: 241 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/570,311

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/647,119

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/241,640

; FILING DATE: 08-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UF15.C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (904) 375-8100

; TELEFAX: (904) 372-5800

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-2

Query Match 100.0%; Score 2641; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.4e-205;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKNSLFSLAIVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60
Db 1 MRKNSLFSLAIVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60

QY 61 ARIILEAHVDWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEKVPVNA 120
Db 61 ARIILEAHVDWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
Db 121 DASFSPTNFVLDGTASADIPAGTYDYVVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

QY 181 VQRQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVLNESFDQ 240
Db 181 VQRQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVLNESFDQ 240

QY 241 TLPNGWTMIDADGDGHNLSTINVYNTATHTDGMFSGKSWTASGGAKIDLSPDNYLVT 300
Db 241 TLPNGWTMIDADGDGHNLSTINVYNTATHTDGMFSGKSWTASGGAKIDLSPDNYLVT 300

QY 301 KVTVPENGKLSYVWSSQVPTWNEHYGVFLSTTGNEAANFTIKLEETLGSXKDPAPMNLVK 360
Db 301 KVTVPENGKLSYVWSSQVPTWNEHYGVFLSTTGNEAANFTIKLEETLGSXKDPAPMNLVK 360

QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGGSNDYTYV 420
Db 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGGSNDYTYV 420

QY 421 YRNVVIAQNLAAATFNQENVAPOQYNYCEVKYTAGVSKVCKDVTVEGSNEFAHVQNL 480
Db 421 YRNVVIAQNLAAATFNQENVAPOQYNYCEVKYTAGVSKVCKDVTVEGSNEFAHVQNL 480

RESULT 2
US-08-353-485-2
Sequence 2, Application US/08353485
Patent No. 5830710
GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Gaylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-485-2

Query Match 100.0%; Score 2641; DB 2; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.4e-205;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKNSLFSLAIVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60
Db 1 MRKNSLFSLAIVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60

QY 61 ARIILEAHVDWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEKVPVNA 120
Db 61 ARIILEAHVDWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
Db 121 DASFSPTNFVLDGTASADIPAGTYDYVVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

QY 181 VQRQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVLNESFDQ 240
Db 181 VQRQPGDAASVVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVLNESFDQ 240

QY 241 TLPNGWTMIDADGDGHNLSTINVYNTATHTDGMFSGKSWTASGGAKIDLSPDNYLVT 300
Db 241 TLPNGWTMIDADGDGHNLSTINVYNTATHTDGMFSGKSWTASGGAKIDLSPDNYLVT 300

QY 301 KVTVPENGKLSYVWSSQVPTWNEHYGVFLSTTGNEAANFTIKLEETLGSXKDPAPMNLVK 360
Db 301 KVTVPENGKLSYVWSSQVPTWNEHYGVFLSTTGNEAANFTIKLEETLGSXKDPAPMNLVK 360

QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGGSNDYTYV 420
Db 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGGSNDYTYV 420

QY 421 YRNVVIAQNLAAATFNQENVAPOQYNYCEVKYTAGVSKVCKDVTVEGSNEFAHVQNL 480
Db 421 YRNVVIAQNLAAATFNQENVAPOQYNYCEVKYTAGVSKVCKDVTVEGSNEFAHVQNL 480

QY 481 TGSVAVGQKVTWKWDAPN 497
Db 481 TGSVAVGQKVTWKWDAPN 497

RESULT 3
US-08-570-311-14
Sequence 14, Application US/08570311
Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann

QY 352 KPAPM----- 356
Db 1086 SPEAMEGR:CGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDELVEIKANGKRADEFTEF 1145
QY 357 -----NIVKS----- 361
Db 1146 ESGTHGEAPAEWTTIDADGDGQWCLSSGQDLWLTAGHTGNVVSFSWNGMALNPDNYL 1205
QY 362 ----- 361
Db 1206 ISKDVGTATKVKYVAVNDGFPDGHVAMISKTGTNAGDFTVVFEETPNKNGGARFGL 1265
QY 362 -----EGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDV--AVSGEGSSND 415
Db 1266 STEADGAKPQSVWIERTVLP--AGTK-YVAFRHYNCSDLNLYLLDDIIOFTMGSGTPTD 1322
QY 416 YIYTVYRDNVVAQNLAATTFOENAVPQYNYCVYKTAGVSPKCKDVTVVEGSGNEFA 475
Db 1323 YTVYVYRDGTKIKEGLTETTFEEDGVATGNHGYCVYKTAGVSPKCKVNTV--NSTQFN 1381
QY 476 HVONLTGSVAGQKVTLLKWDAPN 497
Db 1382 PVKNLKAQPDGDDVVLKWEAPS 1403
RESULT 5
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1
Query Match 35.8%; Score 944.5; DB 4; Length 1477;
Best Local Similarity 31.9%; Pred. No. 1.7e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;
QY 22 TAAAGQ-----GPKTAPSVTHQAVQK-GIR---TSKVLDLRDPIPAGMARIILEAH 68
Db 444 TATTQCKVTLKWDAPSTKTNATTNARSVDGIRELVILSVSDAPELLRSQGAIVLEAH 503
QY 69 DVNEDGTGYQLMDADHNNQYASIPBE--SPFANGTIPAGLYDPPEYKVPVNADASFSP 126
Db 504 DVNWDGSGYQILLDADHDQGVIPSDTHLW-PNCVSPANLPAPEYTVPENADPSCSP 562
QY 127 TNFVLDTASADTPAGTYDYVINPNPGI-IYIVGEGVSKGNDYVVEAGTYHTFTVQROG 185
Db 563 TNMIMDGTASVINIPAGTYDYFAIAAPOANAKIWTAGQGTKEDDYVPEAGKXFLMKMG 622
QY 186 PGDAASW----- 193
Db 623 SGDGTELTISEGGSDYTYVYRDGTKIKEGLTATTTFEEDGVATGNHGYCVYKTAGVS 682
QY 194 -----VTGEGGNEFPAPQNLQMSVSGQTVTLTWQAPAS-----DKRTYVLNESF 237
Db 683 PKVKCDVTVGSGNEFPAPQNLQMSVSGQTVTLTWQAPAS-----DKRTYVLNESF 742
QY 238 DTCTLPNGWTMIDADGDGHNWLTSTINVTATHTGGMFSGKSWTASGGAKIDLSPDNYL 297

Db 743 E-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSNGCVYSESF-CLGGIGV-LTPDNYL 798
QY 298 VTPKTVTPBNGKLSVWSSQ-VPTWNEHYGVFLSTTGNBAANFTIKLLEETLGS----- 350
Db 799 ITPALDLPNGGKLTFWCAQDANYASERYAVYASSTGNDASNTTALLLETTITAKGVRSR 858
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 859 EAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDELVEIKANGKRADEFTEF 918
QY 367 -----PAPYQERTID-----LSAYAGQV----- 385
Db 919 STHGAPAEWTTIDADGDGQWCLSSGQDLWLTAGHTGNVVSFSWNGMALNPDNYL 978
QY 386 ----- 385
Db 979 KDVGTATKVKYVAVNDGFPDGHVAMISKTGTNAGDFTVVFEETPNKNGGARFGL 1038
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTYV 420
Db 1039 EANGAKPQSVWIERTVLPAGTKYVAFRHYNCSDLNLYLLDDIIOFTMGSGTPTD 1098
QY 421 YRDNVVAQNLAATTFOENAVPQYNYCVYKTAGVSPKCKDVTVVEGSGNEFAHVQNL 480
Db 1099 YRDGTKIKEGLTETTFEEDGVATGNHGYCVYKTAGVSPKCKVNTV--NPTQFNPKNL 1157
QY 481 TGSVAGQKVTLLKWDAPN 497
Db 1158 KAQPDGDDVVLKWEAPS 1174
RESULT 6
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1704 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-336-308A-10

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
 Best Local Similarity 31.9%; Pred. No. 2.1e-67;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

22 TAAAGQ-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPDPAGMARILLAH 68
 671 TATTGQKVTWKWDAPSKTNATTWARSVDGIRELVLLSVSDAPELLRSQAELVLEAH 730
 69 DWEDGTGYQLMDADHNGYASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
 731 DVWNDGSGYQLLDADHDQYGVIPSDTHTLW-PNCSPANLFAPEYTVPENADPSCSP 789
 127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185
 790 TNMIMDGTASVNIIPAGTYDYFAIPAQANAKIWIAGQGETKEDDYVFEAGKTYHFLMKKMG 849
 186 PGDAASVV----- 193
 850 SGDGTSLTISEGGSDDYTYVVRDGTIKI EGLTATTFFEDGVATGNHEYCVVEKVTAGVS 909
 194 -----VTGEGNFPAPVQNLQWSVSGQTVLTWQAPAS-----DKRTYVLNESF 237
 910 PKVKDVTVEGSENEFAPVQNLTGSAVGQKVTWKWDAPNGTNPNNPNNPNTTLLSESF 969
 238 DTQTLPLNGWTMIDADGHNWLSNTINVTNTHTGDMGAFKSWTASGAKIDLSPDNYL 297
 970 E-NGIPASWKITDADGDHGW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPDNYL 1025
 298 VTPKVTVPENGLSVYSSQ-VPWTNEHYGVFLSTTGNAAFTIKLLEETLGS----- 350
 1026 ITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDASNTFALLLEETITAKGVRS 1085
 351 -----DKPAPMNLV-----KSGVKL----- 366
 1086 EAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTETPES 1145
 367 ---PAPQERTID-----LSAYAGQV----- 385
 1146 STHGAPAEWTIDADGGQGWHLSSGQDLWLTARHGTNVVASFWSNGMALNPDPNYLIS 1205
 386 ----- 385
 1206 KDVTGATKVKYVAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGARFGLST 1265
 386 -----YLAFFHPSNTGIFRLYLDV--AVSGEGSSNDYTYTV 420
 1266 EANGAKPOSVMIERVLDLPAGTKYVAFRHYGCSOLNYLLDDIQFTMGGSPTDITYTV 1325
 421 YRDNVVIAQNLAAATFNQENAPGOYNCVEKVTAGVSPKVKCDKDTVEGSENEFAHVNLI 480
 1326 YRDGTKI EGTEFTFEEDGVATGNHEYCVVEKVTAGVSPKECVNVTI-NPTQNPVKNL 1384
 481 TGSAVGQKVTWKWDAPN 497
 1385 KAQPDGSDVWLKWEAPS 1401

RESULT 7

US-08-822-324-6

Sequence 6, Application US/08822324

Patent No. 6129917

GENERAL INFORMATION:

APPLICANT: Potempa, Jan S.

APPLICANT: Travis, James

APPLICANT: Genco, Caroline A.

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
 TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
 TITLE OF INVENTION: METHODS

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder

STATE: CO

COUNTRY: US

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/822,324

FILING DATE: 21-MAR-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/013,945

FILING DATE: 22-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 103-95 WO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 488-8080

TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1704 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-822-324-6

Query Match

Best Local Similarity 35.8%; Score 944.5; DB 3; Length 1704;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

22 TAAAGQ-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPDPAGMARILLAH 68
 671 TATTGQKVTWKWDAPSKTNATTWARSVDGIRELVLLSVSDAPELLRSQAELVLEAH 730
 69 DWEDGTGYQLMDADHNGYASIPEE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
 731 DVWNDGSGYQLLDADHDQYGVIPSDTHTLW-PNCSPANLFAPEYTVPENADPSCSP 789
 127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185
 790 TNMIMDGTASVNIIPAGTYDYFAIPAQANAKIWIAGQGETKEDDYVFEAGKTYHFLMKKMG 849
 186 PGDAASVV----- 193
 850 SGDGTSLTISEGGSDDYTYVVRDGTIKI EGLTATTFFEDGVATGNHEYCVVEKVTAGVS 909
 194 -----VTGEGNFPAPVQNLQWSVSGQTVLTWQAPAS-----DKRTYVLNESF 237
 910 PKVKDVTVEGSENEFAPVQNLTGSAVGQKVTWKWDAPNGTNPNNPNNPNTTLLSESF 969
 238 DTQTLPLNGWTMIDADGHNWLSNTINVTNTHTGDMGAFKSWTASGAKIDLSPDNYL 297
 970 E-NGIPASWKITDADGDHGW-KPGNAPGIAGYNSGCVYSEF-GLGGIGV-LTPDNYL 1025
 298 VTPKVTVPENGLSVYSSQ-VPWTNEHYGVFLSTTGNAAFTIKLLEETLGS----- 350
 1026 ITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGNDASNTFALLLEETITAKGVRS 1085
 351 -----DKPAPMNLV-----KSGVKL----- 366
 1086 EAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTETPES 1145

QY 367 ----PAPYQERTID-----LSAVAGQOV----- 385
Db 1146 STHGEAPAEWTTIDADGGGWLCLSSGQDLMTARHGGTNVASFWSNGMALNPDNYLIS 1205
QY 386 ----- 385
Db 1206 KDVTGATKVKYVAVNDGPGDHVAVMISKTGTNAGDFTVVFEETPNGINKGARGFLST 1265
QY 386 -----YLAFRHNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db 1266 EANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNVLDDDIQFTMGGSPTPTDYTYV 1325
QY 421 YRDNVVIQNLAAATTENQENVAPGQYNYCUEVKYTAGVSPKCKDVTEGSGNEFAHVQNL 480
Db 1326 YRDGTIKKEGLTETTFEEDGVATGNHEYCUEVKYTAGVSPKCVNVTI--NPTQFNPKNL 1384
QY 481 TGSVAGQKVTWKWDAPN 497
Db 1385 KAQPDGDDVVLKWEAPS 1401

RESULT 8
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-490-931-10

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
Best Local Similarity 31.9%; Pred. No. 2.1e-67;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;
QY 22 TAAAG-----GPKTAPSVTHQVQK--GIR---TSKYXDLRDPDPAGMARILIEAH 68
Db 671 TATTQCKQVTLKWDAPSTKTNATTARSDVGIRELVLLSVSDAPELLASGQAEIVLEAH 730
QY 69 DWYEDGTGYOMLWDADHNOYGASIPEE--SPFANGTI PAGLYDPEYKVPVNVADASFSP 126
Db 731 DWYDGSQVQILLDADHDQYGVILSDTHLM--PNCVSPANLFAPEYTVVFNADSCSP 789
QY 127 TNFVLDTASADIPAGTYDYVIINPNPGI--IYIVGEGVSKGNDYVVEAGKTYHTVQROG 185
Db 790 TNMIMDGTASVNIIPAGTYDFAIAAPQANAKIWIAGQPTKEDDYVFEAGKCKHFLMKWG 849
QY 186 PGDAASV----- 193
Db 850 SGEDELITISEGGGSDYTYVYRDGTIKKEGLTATTFEEDGVATGNHEYCUEVKYTAGVS 909
QY 194 -----VTGEGNEFAPVONLQWSVSGQVTLTWOAPAS-----DKRTYVLNESF 237
Db 910 PKVCKDVTVEGNEFAPVQNLTGSAVGQKVTILKWDAPNGTNPMPNPNGTTLSESF 969
QY 238 DTQTLPGNWTMIDADGDGHNWLTSTINVTATHTGDMFSPKSWTASGGAKIDLSPDNYL 297
Db 970 E-NGIPASWKTIDADGDGHG--KPGNAPGIAGYNSNGCVVSSEF--GLGIGV--LTPDNYL 1025
QY 298 VTPKVTVPENGLSVWVSQ--VPWTNEHYGVFLSTTGNEAANFTIKLLETLGS----- 350
Db 1026 ITPALDLPNGGKLTFFWVCAQDANYASEHYAVYVASSGTNDASNFTNALLBETITAKVRSR 1085
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 1086 EAIRGRIQGTWRQKTVDLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFS 1145
QY 367 ----PAPYQERTID-----LSAVAGQOV----- 385
Db 1146 STHGEAPAEWTTIDADGGGWLCLSSGQDLMTARHGGTNVASFWSNGMALNPDNYLIS 1205
QY 386 ----- 385
Db 1206 KDVTGATKVKYVAVNDGPGDHVAVMISKTGTNAGDFTVVFEETPNGINKGARGFLST 1265
QY 386 -----YLAFRHNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db 1266 EANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNVLDDDIQFTMGGSPTPTDYTYV 1325
QY 421 YRDNVVIQNLAAATTENQENVAPGQYNYCUEVKYTAGVSPKCKDVTEGSGNEFAHVQNL 480
Db 1326 YRDGTIKKEGLTETTFEEDGVATGNHEYCUEVKYTAGVSPKCVNVTI--NPTQFNPKNL 1384
QY 481 TGSVAGQKVTWKWDAPN 497
Db 1385 KAQPDGDDVVLKWEAPS 1401

RESULT 9
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted M. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville

```
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 amino acids

QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
D5 1127 ESSTHGEAPAEWTTIDADGCGQWCLSSGQLDMLTAHGGTNVNVASFWSWNGMALNPDNVL 1186
QY 386 ----- 385
D5 1187 LSKDVTGATKVYVAVNDGPGDHYAVMISKTGTNAGDFTVVFETPNGINKGARGFL 1246
QY 386 -----YLAFRFNSTGIFRLYLDVV--AVSGEGSSNDYTY 418
D5 1247 STEANGAKPQSVWIERVTVDLPAGTKYVAFRHYNCSDLYILLDDIQTMGGSPTDITY 1306
QY 419 TVYRDNVVAQNLAAATFENQENVAPGQVNYCCEVKYTAGVSPKVCCKDVTVEGSEFAHVQ 478
D5 1307 TVYRDTKIKESGLTETTFEEDGVATGNHEYCCEVKYTAGVSPKVCVNTI-NPTQFNPK 1365
QY 479 NLTSVAVGQKVTWKWDAPN 497
D5 1366 NLKAQPDGDDVVLKWEAPS 1384

RESULT 10
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Arn
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1358 amino acids
```

```
Query Match 35.7%; Score 943.5; DB 2; Length 1687;
Best Local Similarity 31.8%; Pred. No. 2.5e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVXDLRDPDIPAGMARIILRAH 68
D5 652 TATTGQKVTWKWDAPSTKTATNTARSVDGIRELVLLSVDAPELLRSQAEIVLEAH 711
QY 69 DWEDGTGYMLADHNOYGASIEPE--SFWFANGTIPAGLYDPPEYKVPVNADASP 126
D5 712 DVWNDGSGYQILLADHDQYGGVIPSDDTHLW-PNCSPANLFAPEYTVPENADPSCSP 770
QY 127 TNFVLDGTASADIPAGTYDYVIINPQGI-IYIVGEGSVKGNDDYVEAGKTVHFTVQROG 185
D5 771 TNMINDGTASVNIPTAGTYDFAIAAPQANAKIWAQGGPTKEDDYVEAGKTVHFLMKWG 830
QY 186 PGDAASVV----- 193
D5 831 SGDGPTELATSEGGSDYTYVYVRDGTIKI EGLTATTTFFEDGVATGNHEYCEVKYTAGVS 890
QY 194 -----VTGEGNEFAPQNLQWSVGGTTLTWQAPAS-----DKRTYVINE 235
D5 891 PKVKCDVTVEGSENEFAPQNLGTSAVGQKVTWKWDAPNGTTPNPNPNPNPGTTTISE 950
QY 236 SPDTQTLNGWTMIDADGDHNLSTINVTATHTGCGAMPKSKWTSAGGAKIDLSPDN 295
D5 951 SFE-NGIPASWKTIDADGDHGW-KFGNAPGTAGYNSGCVYSESF-GLGGIGV-LTPDN 1006
QY 296 YLVTPKVTPVNGKLSYVSSQ-VPWNTNEHYGVFLSTTGNEAANTFKILEFTLAS---- 350
D5 1007 YLITPDLPLNGKGLTFWCAQDANYASEHYAVYASSTGNDASNTNALLEETITAKGVR 1066
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
D5 1067 SPEAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLIDVEIKANGKRAADFTTF 1126
```



```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-27

Query Match      30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%; Pred. No. 3.7e-57;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKRTSKVKDLRD-----PIPGAMARILEA 67
Db 328 TATTQGGKVTLKWDAPS-----AKKAESREVKRIKIGDLFVTTEPANDVRANEKVVLA 382
QY 68 HDVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDP-PEYKVPVNADASFSP 126
Db 383 DNWAGDNTGYQFLDADHNTFGSVIPATGFLF-TGTASSNLYSANFEYLVPANADPVVTT 441
QY 127 TNFVLDTGASADIPAGTYDYVIINPNP--GIIVYVGEV---VSKGNDYVVEAGKTYHFTV 181
Db 442 QNIIVTGGQEVVIPGGVYDYCIITNPEPASCKMWIAGDGNOPARYDDTFEAGKKTFTM 501
QY 182 QROGPGDAASVV----- 193
Db 502 RRAGMGDGTDMVEDDPSASYTYTVYRDGTKIKEGLTATTTFEEDGVAAGNHEVCVEVKT 561
QY 194 -----VTGEGNEFAPQNLQWSVSGQTVTLTWOAPAS-----DKRTVVLNE 235
Db 562 AGVSPKVKDVTVEGSENEFAPVQNLGTSGAVGQKVTLLKWDAPNGTNPENPFGTTLSE 621
QY 236 SPTQTLPLNGWTMIDADGDNHNLSTINVTATHTGDMGAFMSKSWTASGAKIDLSPDN 295
Db 622 SFE-NGIPASWKTIDADGDGHWG-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTFPDN 677
QY 296 YLYTPKVTVPENGLSWVSSQ-VPWTNHYGVFLSTTGNEAANFTIKLLEETLS----- 350
Db 678 YLITPDLALANGKLTFWCAQDANYASEHYAVYASSTGNDASNTNALLEETITAKVGR 737
QY 351 -----DKPAPMNLV-----KSRGVKL----- 366
Db 738 SPEAIRGRIQGTWRQKTVDLIPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADETTF 797
QY 367 -----PAPQERTID-----LSNAGQQV----- 385
Db 798 ESSTHGEAPAEWTITADGDGQDWLCLSSGQLDWLTAHGQTNVVASFSWNGMALNPDNYL 857
QY 386 ----- 385
Db 858 ISKDVGTATKVKYAYAVNDGFGPDGHYAVMISKGTNAGDFTVVFETPNINKGARFGL 917
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTY 418
Db 918 STEANGAKPQSVLIERIVDLIPAGTKYVAFRHYNCSLDLYILLDDIQFTMGSPPTPDYTY 977
QY 419 TVYRDNVVIAQNLAAATTFNGENVAPQYNTCVVEKYTAGSVKPKVDVTVEGSENEFAHVQ 478
Db 978 TVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGSVKPKVCNVNVTI-NPTQFNPK 1036
QY 479 NLGSAVGQKVTLLKWDAPN 497
Db 1037 NLXAPDGGDVVLKWEAPS 1055

RESULT 11
US-08-570-311-10
; Sequence 10, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proculus-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph

```

```

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 03-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UP15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1732 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-10

Query Match      30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1e-56;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKRTSKVKDLRD-----PIPGAMARILEA 67
Db 693 TATTQGGKVTLKWEAPS-----AKKAESREVKRIKIGDLFVTTEPANDVRANEKVVLA 747
QY 68 HDVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDP-PEYKVPVNADASFSP 126
Db 748 DNWAGDNTGYQFLDADHNTFGSVIPATGFLF-TGTASSNLYSANFEYLVPANADPVVTT 806
QY 127 TNFVLDTGASADIPAGTYDYVIINPNP--GIIVYVGEV---VSKGNDYVVEAGKTYHFTV 181
Db 807 QNIIVTGGQEVVIPGGVYDYCIITNPEPASCKMWIAGDGNQPARVDDTFEAGKKTFTM 866
QY 182 QROGPGDAASVV----- 193
Db 867 RRAGMGDGTDMVEDDPSASYTYTVYRDGTKIKEGLTATTTFEEDGVAAGNHEVCVEVKT 926
QY 194 -----VTGEGNEFAPQNLQWSVSGQTVTLTWOAP-----ADKRTYVLN 234
Db 927 AGVSPKVKDVTVEGSENEFAPVQNLGTSGVQKVTLLKWDAPNGTNPENPFGTTLSE 986
QY 235 SPTQTLPLNGWTMIDADGDNHNLSTINVTATHTGDMGAFMSKSWTASGAKIDLSPD 294
Db 987 ESFE-NGIPASWKTIDADGDGHWG-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTFD 1042

```


APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSER: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL USA
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1087 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-8

Query Match 30.5%; Score 806.5; DB 2; Length 1087;
Best Local Similarity 29.4%; Pred. No. 1.6e-56;
Matches 211; Conservative 76; Mismatches 166; Indels 265; Gaps 19;

QY 39 AVOKGRTSKVKDLRD-----PIPGAMARIILEAHVDWEDGTGYQMLWDADHNOY 88
Db 73 SAKKAASREVKRIGDLFTVIEPANDVRANEAKVLAADNVWGDNTGYQLLDADHNTF 132
QY 89 GASIPESRFANGTIPAGLYDP-FEYKVPVNADASFSPNFVLDTGASADIPAGTYDYV 147
Db 133 GSVIPATGPIF-TGTASSNLYSANFEVLIIPANADPVVTTQNIIVTGGQGVWIPGVGYDYC 191
QY 148 IINPNP-GIIYIVGEG---VSKGNDYVVEAGKTYHFTVQRQPGDAASVW----- 193
Db 192 ITNPEPASGKOWIAGDGDNDQPARYDDFTPEAGKKYFTMRAGMGDGTDMVEEDDPSAY 251
QY 194 -----VTGGGSGNEFAP 204
Db 252 TYTVYRDGTIKKGLTATTTEEDGVAAGNHEVCVEVKYTAGVSPKVKDVTVEGSENEFAP 311
QY 205 VQNLQMSVSGQTVTLTWOAPAS-----DKRTYVLNESFDQTLPNGHTMTIDADGDGH 256
Db 312 VQNLGSAVGQKVLTKWDAPNGTPNPNPNPGTITLSESEFE-NGIPASMKITIDADGDGH 370
QY 257 NWLSTINVTATHTGDMGAFSKSWTASGAKIDLSPDNLIYTPKVTVPENGLSYWVSS 316

Db 371 GW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPDNYLITPALDLANGGKLTFFVCA 427
QY 317 Q-VPWNHGYGVFLSTTGTNEAANFTIKLEETLGS-----DKP 353
Db 428 QANTASESHYAVYASSTGNDASFTNALLEETITAKGVRSPKAIKRGITQGTWRQKTVDL 487
QY 354 APMNLV-----KSEGVKL-----PAPYQERTID----- 376
Db 488 AGTKYVAFPHFQSTDMFYIDLDEVEIKANGKREADFTETPESSTHGEAPAEWTTIDADGDG 547
QY 377 -----LSAYAGQGV----- 385
Db 548 QDWLCLSSGQLDMLTAHGTTNVASFWSNGMALNPDNYLISKDVTGATKVKYYAVNDGF 607
QY 386 ----- 385
Db 608 PGDHYAVMISKGTGNAGDFTVVFETPNNGINKGARFGLSTEANGAKPQSVWIERTVDLP 667
QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGGSSNDYTYTVYRDNVVIAQNLAAATFNQE 439
Db 668 AGTKYVAFPHFVNCSDLDVILLDDIQFTMGGSPTDPTDYTVVYRDGKIKGLTETTFEED 727
QY 440 NVAPQYQYNYCYEYKYTAGVSPKVKCDVTVEGSENEFAHQNLGTSVAGQKVLTKWDAPN 497
Db 728 GVATGNHEVCVEVKYTAGVSPKVCVNVITI-NPTQPNFVKLKAQPDGDDVVLKWEAPS 784

RESULT 14
US-08-353-485-8
Sequence 8, Application US/08353485
Patent No. 5830710
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tumwasorn, Somving
APPLICANT: Lapine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:32:49 ; Search time 19.8443 Seconds

(without alignments)
6969.043 Million cell updates/sec

Title: US-08-353-485-2

Perfect score: 2641

Sequence: 1 MRKLSRSLAVLLSLCWG.....QLTGSAVGQVTLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 951.5 | 36.0 | 1706 | 14 | US-10-229-066-10 |
| 2 | 806.5 | 30.5 | 1732 | 14 | US-10-229-066-11 |
| 3 | 791.5 | 30.0 | 419 | 15 | US-10-174-695-3 |
| 4 | 662.5 | 25.1 | 419 | 15 | US-10-174-695-5 |
| 5 | 457.5 | 17.3 | 231 | 15 | US-10-174-695-6 |
| 6 | 188.5 | 7.1 | 196 | 15 | US-10-174-695-4 |
| 7 | 150 | 5.7 | 29 | 15 | US-10-387-977-27 |
| 8 | 150 | 5.7 | 1483 | 12 | US-10-282-122A-51483 |
| 9 | 148 | 5.6 | 42 | 14 | US-10-229-066-3 |
| 10 | 145 | 5.5 | 509 | 15 | US-10-387-977-101 |
| 11 | 144 | 5.5 | 31 | 15 | US-10-387-977-6 |
| 12 | 146 | 5.4 | 1946 | 12 | US-10-282-122A-62947 |
| 13 | 141 | 5.3 | 31 | 15 | US-10-387-977-9 |
| 14 | 139 | 5.3 | 26 | 15 | US-10-387-977-15 |
| 15 | 138.5 | 5.2 | 2468 | 12 | US-10-282-122A-66335 |

| | | | | | | |
|----|-------|-----|------|----|----------------------|-------------------|
| 16 | 138.5 | 5.2 | 2468 | 14 | US-10-246-330-4 | Sequence 4, Appl |
| 17 | 137 | 5.2 | 25 | 15 | US-10-387-977-82 | Sequence 82, Appl |
| 18 | 137 | 5.2 | 579 | 15 | US-10-369-493-9075 | Sequence 9075, Ap |
| 19 | 134 | 5.1 | 2435 | 12 | US-10-282-122A-47453 | Sequence 47453, A |
| 20 | 132 | 5.0 | 698 | 14 | US-10-156-761-11953 | Sequence 11953, A |
| 21 | 129 | 4.9 | 973 | 14 | US-10-156-761-9394 | Sequence 9394, Ap |
| 22 | 128 | 4.8 | 26 | 15 | US-10-387-977-14 | Sequence 14, Appl |
| 23 | 127 | 4.8 | 811 | 12 | US-10-272-898-7 | Sequence 7, Appl |
| 24 | 127 | 4.8 | 811 | 14 | US-10-011-366-7 | Sequence 7, Appl |
| 25 | 127 | 4.8 | 811 | 15 | US-10-354-774-7 | Sequence 7, Appl |
| 26 | 127 | 4.8 | 811 | 15 | US-10-271-012-7 | Sequence 7, Appl |
| 27 | 127 | 4.8 | 2710 | 12 | US-10-272-898-6 | Sequence 6, Appl |
| 28 | 127 | 4.8 | 2710 | 14 | US-10-011-366-6 | Sequence 6, Appl |
| 29 | 127 | 4.8 | 2710 | 15 | US-10-354-774-6 | Sequence 6, Appl |
| 30 | 127 | 4.8 | 2710 | 15 | US-10-271-012-6 | Sequence 2, Appl |
| 31 | 126 | 4.8 | 1752 | 14 | US-10-387-388-2 | Sequence 21739, A |
| 32 | 125.5 | 4.8 | 577 | 15 | US-10-369-493-21739 | Sequence 21739, A |
| 33 | 124 | 4.7 | 2013 | 12 | US-10-282-122A-60608 | Sequence 60608, A |
| 34 | 124 | 4.7 | 2358 | 12 | US-10-282-122A-45763 | Sequence 45763, A |
| 35 | 123.5 | 4.7 | 1204 | 12 | US-10-282-122A-49627 | Sequence 49627, A |
| 36 | 123 | 4.7 | 866 | 14 | US-10-222-038-2 | Sequence 2, Appl |
| 37 | 123 | 4.7 | 951 | 9 | US-09-924-097-15 | Sequence 15, Appl |
| 38 | 122 | 4.6 | 806 | 15 | US-10-369-493-3678 | Sequence 3678, Ap |
| 39 | 121 | 4.6 | 25 | 9 | US-09-847-185-46 | Sequence 46, Appl |
| 40 | 121 | 4.6 | 25 | 10 | US-09-930-915A-18 | Sequence 18, Appl |
| 41 | 121 | 4.6 | 25 | 14 | US-10-224-286-46 | Sequence 46, Appl |
| 42 | 121 | 4.6 | 25 | 14 | US-10-082-014-40 | Sequence 40, Appl |
| 43 | 121 | 4.6 | 25 | 14 | US-10-372-076-41 | Sequence 41, Appl |
| 44 | 121 | 4.6 | 25 | 15 | US-10-387-977-76 | Sequence 76, Appl |
| 45 | 121 | 4.6 | 25 | 15 | US-10-387-977-79 | Sequence 79, Appl |

ALIGNMENTS

RESULT 1

US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 36.0%; Score 951.5; DB 14; Length 1706;
Best Local Similarity 31.9%; Pred. No. 3.5e-75;
Matches 237; Conservative 74; Mismatches 156; Indels 275; Gaps 20;
QY 22 TAAAG-----GPKTAPSVTHQAVK--GIR---TSKVLDLDPDPAGVARIIEAH 68
Db 671 TATTQCKVTLKWDAPSTKTNTATTARSVDGIRIELVLSVSDAPELLRSQAIVLEAH 730
QY 69 DWYEDGTGYOMLWDADHNOYGASIPDE--SFFANGTIPAGLYDPEYKVPVNADASFSP 126
Db 731 DWVNDGSGYQILLDDADHDQIGQVIFSDTTLW-PNCSPANLFAPEYTVPNADPSCSP 789

```

QY 127 TNEVLDTASADIPACTYDYVIINPFI--IYIVGVSXGNDYVYVAGKYVHTVQRC 185
Db 790 TNMIMDGTASVNIPTAGTYFAIAAPQANAKIWIAGQPTKEDDYVFEAGKXHFLLMKRMG 849
QY 186 PGDAASV----- 193
Db 850 SGGTELTISEGGSDYTYTVYRDGTKIKEGLTATTFFEDGVATGNHEYCVVKYTAGVS 909
QY 194 -----VTGEGNEFAPVQNLQWSVSGOTVTLTWQAPAS-----DKRTYVLNE 235
Db 910 PKVKCDVTVEGSENEFAPVQNLQWSVSGOTVTLTWQAPAS-----DKRTYVLNE 235
QY 236 SPDTQTLNGWTMIDADGGHNLSTINVTATHTGDMRFSKSWTASGAKIDLSPDN 295
Db 970 SFE-NGIPASWKTIADGGHGW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPD 1025
QY 296 YLVTPTKVPENGLKSYWSSQ-VPTWNEHYGVFLSTTGNAAANFTIKLEETL---GSD 351
Db 1026 YLITPALDLNGGKIPTWVCAQDANYASEHYAVYASSTGNDASFTNALLEETITAKGVR 1085
QY 352 KPAPM----- 356
Db 1086 SPEAMGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTF 1145
QY 357 -----NLVKS----- 361
Db 1146 ESSTHGEAPAEWTTIDADGGQGWCLSSQGLDMLTAHGTTNNVSSPSWNGMALNPNDYL 1205
QY 362 ----- 361
Db 1206 ISKDVGTATKVKYVAVNDGPPGDBHYAVMISKTGNAGDFTVFEETPNKNGGARFGL 1265
QY 362 -----EGVKLPAPQERIDISAVAGQOVLAFRHFNSTGIFRLYLDV--AVSSEGSND 415
Db 1266 STEADGAKPOSVMIERVTLDP--AGTK-YVAFRHYNCSDLNILLDDIQTWGGSPPTD 1322
QY 416 YTVYRDGVTKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSPKCKDVTVGSGNEFA 475
Db 1323 YTVYRDGVTKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSPKCKDVTVV-NSTQFN 1381
QY 476 HVQNLTGSVAGQKVTWKWDAPN 497
Db 1382 PVKNLKAQPGDGVVLKWEAPS 1403

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

```

Query Match 30.5%; Score 806.5; DB 14; Length 1732;
 Best Local Similarity 29.7%; Pred. No. 3.2e-62;

```

Matches 220; Conservative 72; Mismatches 172; Indels 277; Gaps 22;
QY 22 TAAAGGPKT-----APSVTHQAVQKIRTSKVKDLRD-----PIPAQMARILLEA 67
Db 693 TATTQSQKVTLKWEAPS-----AKKAEGSRVKRIGDGLFVIEPANDVRANEAKVVLAA 747
QY 68 HDVWEDGTGYQLMDADHNOYGASIPESFWEFANGTIPAGLYDP-PEYKVPVNVNADASFSP 126
Db 748 DNVWGDNTGYQLMDADHNTFSGVIPATGFLP-TGTASNLSYANFEYILIPANADPVVIT 806
QY 127 TNFVLDTASADIPACTYDYVIINPFI--GIYIVGEG--VSKGNDYVVEAGKYVHTV 181
Db 807 QNLIVTQCEVWIPGGVYDYCITNPEPASGKMWIAGDGNQAPARYDDFTFEAGKYTFM 866
QY 182 QROGPDAAASV----- 193
Db 867 RRAGMGDGTMEVEDDSRASYTYTVYRDGTKIKEGLTATTFFEDGVAAAGNHEYCVVKY 926
QY 194 -----VTGEGNEFAPVQNLQWSVSGOTVTLTWQAP-----ASDKRTYVLN 234
Db 927 AGVSPKCKDVTVGSENEFAPVQNLQWSVSGOTVTLTWQAP-----ASDKRTYVLN 234
QY 235 ESPTDTLNGWTMIDADGGHNLSTINVTATHTGDMRFSKSWTASGAKIDLSPD 284
Db 987 ESPE-NGIPASWKTIADGGHGW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPD 1042
QY 295 NYLVTPTKVPENGLKSYWSSQ-VPTWNEHYGVFLSTTGNAAANFTIKLEETLGS--- 350
Db 1043 NYLITPALDLNGGKIPTWVCAQDANYASEHYAVYASSTGNDASFTNALLEETITAKGV 1102
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 1103 RSPKALRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTF 1162
QY 367 -----PAPYQERTID-----LSAVAGQOV----- 385
Db 1163 FESSTHGEAPAEWTTIDADGGQGWCLSSQGLDMLTAHGSSNVVSSFSWNGMALNPNDY 1222
QY 386 ----- 385
Db 1223 LISKDVGTATKVKYVAVNDGPPGDBHYAVMISKTGNAGDFTVFEETPNKNGGARFG 1282
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSSEGSNDYT 417
Db 1283 LSTEANGAKPOSVMIERVTLDPAGTKYVAFRHYNCSDLNILLDDIQTWGGSPPTD 1342
QY 418 YTVYRDGVTKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSPKCKDVTVGSGNEFA 477
Db 1343 YTVYRDGVTKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSPKCKDVTVV-NSTQFN 1401
QY 478 QNLTSASV--GQKVTWKWDAP 496
Db 1402 QNLTAQAPNSMDAILKNAP 1422

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8

```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match          30.0%; Score 791.5; DB 15; Length 419;
Best Local Similarity 39.9%; Pred. No. 8e-62;
Matches 174; Conservative 61; Mismatches 100; Indels 101; Gaps 14;

QY 58 AGMARIILEAHVDVWEDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDP-FEYKV 116
Db 1 ANEAKVVLADADNVWGDNTGYQLLDADHNTGSGVIPATGPLF-TGTASSNLYSANFEYLI 59
QY 117 PVNADASPSPTNFVLDTGASADIPAGTYDYVLIINPNP--GIYIVGEG---VSKGNDYV 171
Db 60 PANADPVVTTQNIIVTGGEVVIGGVYDICTNPPEASGKMWIAGDGGNQPARYDDFTF 119
QY 172 EAGKTYHTVORQPGDAASVV----- 193
Db 120 EAGKTYHTMRRAGMGDGTMEVEDDSPAITYTYVYRDGTKIKEGLTATTTEEDGVAAGN 179
QY 194 -----VTGEGNEFAPVQNIQWSVSGTIVLTWOAP----- 224
Db 180 HEYCEVVKYTAGVSPKCKDVTEGSENEFAPVQNLTGSSVGQKVTWKWDAPNGTNP 239
QY 225 ASDKRTYVLNESFDQTLPNGWMTMDADGGHNLSTINVTATHTGDMGAMFSKSWTAS 284
Db 240 PNPNGPTTLESFE-NGIPASMKTTIDAGDGGHGW-KPGNAPGIAGYNSGCVYIS 296
QY 285 GKAKDLSPDNLYTPKVTVPENGKLSYVWSSQ-VPWTHYGVFLSTTGNEAANFTIKL 343
Db 297 GGIGV-LTPDNLYLIPALDLPNGGKLTFWVCAQDANYASEHYAVASSTGNDASFT 355
QY 344 LEETLGSKDPAMNLVKSEGVKLP-----APYQERTIDLSAYAGQOQVYLAFRHN 395
Db 356 LEET-----ITAKGVRSPEAIRGRIQGTWRQKTVDLP--AGTK-YVAFRHF 401
QY 396 GIFRLYLDVAVSGEG 411
Db 402 DMFYIDLDEVEIKANG 417

RESULT 5
US-10-174-695-6
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-6

Query Match          17.3%; Score 457.5; DB 15; Length 231;
Best Local Similarity 27.8%; Pred. No. 2.2e-32;
Matches 118; Conservative 22; Mismatches 63; Indels 22; Gaps 6;

QY 80 LWDADHNOYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFPTNFVLDTGASAD 138
Db 2 LLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLI PANADPVVTTQNIIVTGG 60
QY 139 IPAGTYDYVLIINPNP--GIYIVGEG---VSKGNDYVVEAGKTYHTVORQPGDAASV 193
Db 61 IPGGVYDYCITNPPEASGKMWIAGDGGNQPARYDDFTFEAGKTYHTVORQPGDAASV 120
QY 194 VTGEGNEFAPVQNIQWSVSGTIVLTWOAPASDKRTYVLNESFDQTLPNGWMTMDAD 253
Db 121 VE----- 122

Query Match          25.1%; Score 662.5; DB 15; Length 419;
Best Local Similarity 36.0%; Pred. No. 2.7e-50;
Matches 157; Conservative 60; Mismatches 118; Indels 101; Gaps 15;
```

INFORMATION: Eric Charles Reynolds.
; PRIORITY DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.


```
; LENGTH: 1483
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51483

Query Match          5.7%; Score 150; DB 12; Length 1483;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 127; Conservative 81; Mismatches 213; Indels 208; Gaps 29;

QY 27 GPKTAP-----SVTHQAVQKGIKRTKVDLRDPIIPAGMARIIIEAHVDWEDGTG- 77
DB 522 GTFVTPSQINVEQGSATDPVKIDLNGNTLKDWDQ-----SGKTLVQGTDTYVTDGIT 577
QY 78 -----QMLWDAHDHNOVGAS-----IPEESPFWANGTIPAGLXDPFEKVPV 118
DB 578 LSQSYLAGLALGQYLTLDFNGGASQTITINVVKNETVKLSVGTIVSGNPGD--TVKVPV 635
QY 119 N-----ADASFPTNF-VLDGTASADIPAGT--YDYVIINPNFGIIYIV----- 159
DB 636 TISQVSTPVGLICMDISYDASKFTVKDVLPTDLVKDTONYSFIVNTSTPGKISITFTDP 695
QY 160 -----GEGVSKGNDYVVEAGKT---YHFTVQROGPDGAASVVVTGEGNEFAPVQNL 208
DB 696 TLANYPSVDGLAYLDFIINSNATAGDSALTY-----DPATLIIVADEND-----KDI 743
QY 209 QMSVSGQTTLTWOAPASDKRTVVLNESFT--QTLF-----NGWTMIDA-DGDGH 256
DB 744 KDAASNGKITVTGSAV--VQSVVNTSSVTYDQNAPOQAVSITFNGVTKDVKDQASG- 800
QY 257 NMLSTINVTANTHTGDGAMFSKSWTASGA-----KIDLSPN-----YLVTPTKVTP 306
DB 801 ---NLTAKGSDVTATSDGITLSQSYLATLAAGTYVTYIDFSAGNAGTFTVWVGKTVVGS 857
QY 307 NGKLSWTVSS-----QVFWT-----NEHYGFLSTTGNEAANFTIK-----LLET- 347
DB 858 ATTAVGTVSGKAGDVTVKPVTISKVTTTPVGLTCABEIDYDASKFTVKDLPNTDLVKD 917
QY 348 ----LGSKPKAPMNL-----VKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRH 391
DB 918 NYSFIVNTSTPGKISITFTDPLANTYPIASDGI---LAYLDFIINSNATAGDSALT 971
QY 392 FNSTGIFRLYLDVAVSGEGSSNDY-----KIDLSPN-----YLVTPTKVTP 417
DB 972 -NPSGFIADENDKDIQDAASNGKITVTGSTPVAENSVVNTSSVTYDQNAPOQAVSITL 1030
QY 418 -----YTVYEDNVVIAQNLAAATFENQENAVAPQVNYHCVEVKY-T 455
DB 1031 NGNTITDVKDASGNTLKAGSDYTVTSDGITLSQSYLAT-----LAAGTYTYTVDF 1084
QY 456 AGVSPKVKCDVTVEGSENEFAHVQNLGSA 484
DB 1085 AGTFTVVAKTVVSSATLAVGTVSGKA 1113

RESULT 9
US-10-229-066-3
; Sequence 3, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-3

Query Match          5.6%; Score 148; DB 14; Length 42;
Best Local Similarity 64.1%; Pred. No. 7.3e-06;
Matches 25; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 58 AGMARIIIEAHVDWEDGTGYQMLDADHNOYGASIPES 96
DB 1 SQGAIVLEAHVDWVWDSGYQILLDADHDQYGVIPSDT 39

RESULT 10
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match          5.5%; Score 145; DB 15; Length 509;
Best Local Similarity 19.1%; Pred. No. 0.00061;
Matches 102; Conservative 65; Mismatches 155; Indels 212; Gaps 22;

QY 114 YKVPVN---ADASF-----SPTNFVLD-----GTASADIPAGTY----D 145
DB 10 YNTPVRLVAVAGKPKALKPWLTKAOKGFYLDVHYTDEAEVGTTNASIKAFIHKYND 69
QY 146 YVIINPNPGIIVIGEGVSKGNDYVVEAGKTYHFTVQROGPDGAASVVVTGEGNEFAPV 205
DB 70 GLAASAPVFLALVGD-----TDVISGEKGTKVKV 100
QY 206 QNLQWS-VSG---QTVTLTWOAPASDKETVYVNE--SFDTOTLPN----GWTMIDADGD 254
DB 101 TDLIYSAVDGDYFPEWYTFRMSASSPEELTNIIDKVLMEKATMPDKSVLEKVLLAGAD 160
QY 255 GHNWLSSTI-----NVYN-----TATHTGDGAMFSKSWTASG 286
DB 161 -YSWNSOVQOPTIKYGMQVYVNOEHGYTDVYNLKPATGYCYSHLNTGVSPA-NYTAHGS 218
QY 287 A-----KIDLSPNYLVTPK-----VTVPPENGKLSYVW 314
DB 219 ETAWADPLLTTSQKALTNKDKYFLAIGNCCITAOFDYVQPCFGEVITRVKKGAYATIG 278
QY 315 SSOVPWTEHY-----GVFLSTTGNEAANFTIKLEET----- 347
DB 279 SPSNSYWGEDIYWSVCANAVFGVQPTFEGTSMGSDATFLEDSYNTVNSIMMAGNLAATH 338
QY 348 -----LGSKPKAPMNL-----LVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRH 383
```

```
Db 339 AGNIGNITHGAYHWEAVHVGDSVMPYRAMPKNTYITLPSLQNCASVSIQASGS 398
QY 384 QVYLAFRHFNSTGIFRLYLDVAVSGEGSNDYTYTVYRDNVVIQNLAAATTNQNENAP 443
Db 399 Y-----VAISKDG-----VLYGTGVANAGSVATVSMTKQITEN 431
QY 444 GOVNYCUEKVTAGVSPKCKDVTVEGSNEFAHVQNLTGSAGVQKVTWKWDAPN 497
Db 432 GNY-----DVVITRSNLYLVKIQVGEPSYQPVSNLTATTQCKVTLKWEAPS 481

RESULT 11
US-10-387-977-6
; Sequence 6, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 05/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-6

Query Match 5.5%; Score 144; DB 15; Length 31;
Best Local Similarity 96.6%; Pred. No. 1e-05;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 469 EGSNEFAHVQNLTGSAGVQKVTWKWDAPN 497
Db 1 EGSNEFAPVQNLTGSAGVQKVTWKWDAPN 29

RESULT 12
US-10-282-122A-62947
; Sequence 62947, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62947
; LENGTH: 1946
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-62947

Query Match 5.4%; Score 143; DB 12; Length 1946;
Best Local Similarity 18.4%; Pred. No. 0.0071;
Matches 110; Conservative 63; Mismatches 214; Indels 210; Gaps 23;

QY 22 TAAAOGGPKTAPSVTHQAVQKGIKRTSKVKDLRD----PIPAQMARIILEAHWDWEDGTG 77
Db 417 TATLLAGGLTVKDIHKKIQVSADGKFTDMNGGTSAPVAAGTTRIT-----RTIGF 469
QY 78 QMLWDADHNOYCASIPESFWFANGTIPAGLIDPPEYKVPVNADASFSTPNFVLGD--TAS 136
Db 470 A-----KEGTVDDNSFYLDNAKLKVG-----NVEINKDTGINAGNQKITGLTDG 514
QY 137 ADIPAGTYDYVIINPNPGIYIVGEGVSKGNDVVVEAGKTYHTVQROGPGDAASVVVVG 196
Db 515 ADDADAVTIKQKLNKAPNL--TPGNGIEINN-----TNSLSDVA 551
QY 197 EGSNEFAPVQNLTGSAGVQKVTWKWDAPN-----SDKRTY-----VLNESFD 238
Db 552 ANGNVTTP-----SYTIGVKITTLTNNGPSDKFAVKDSNTHNSLVTAOKLAGYLNEVNR 607
QY 239 T-OTLENGWTMDADGDGHNLSTINVTATHTGDCAMPSKSWTASGGAKIDLSPDNYL 297
Db 608 TADSALQSFVTVGGDAASNNITLTKTSLSLVAGDNGITVKTDTTQKKVTVGIDQANGL 667
QY 298 VTPKVTVPENGLSVYVWSQVWNEHYGVFLSTTGNEAANFTIKLLEETL-----348
Db 668 TTPKLTVGSNSKTQVIEQV-----IGNDTKNI--IKGLSSTLTDITNDNT 713
QY 349 -----GSDKPA-----PMNLVSKSEGKLPAPYQ 371
Db 714 HTTEQDNDAQKESNAASIKDVLNAGFNQNGKPVDFVSTVDTVNFANGDGTATVTV- 772
QY 372 ERTIDLSAYAGQVYLAFRHN-----STGIFRLYLDVAV-----SGEGSSNDYTYTV 421
Db 773 -----NYNEGKETSTVNVNVNVDVDDTHLANNNGNKKSVKTTTLT 813
QY 422 RDNVIAQNLAAATTF-----NOENVAPGQYNYCVKVTAGVSP-----KVCKD- 465
Db 814 KTN--GANGNATKFSANNNGDALVNAKGIADNLNLAEBEIHHTKGTADTALQTFKVKKG 870
QY 466 -----VTVEGSNEFA-----HVQNLTGSAGVQKVTWKWD 494
Db 871 ATNDDDTITVGNKNAVDTLAFKGENGLTVATKKDGTVTFTGINTQSGLKAGDNTTLNKD 927

RESULT 13
US-10-387-977-9
```

```

; Sequence 9, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-9

Query Match
Best Local Similarity 5.3%; Score 141; DB 15; Length 31;
Matches 27; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 469 EGSNEFAHVQNLTGSAVGQKVLKWDAPN 497
DB 1 EGSNEFAPVQNLTGSSVGQKVLKWDAPN 29

RESULT 14
US-10-387-977-15
; Sequence 15, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-15

Query Match
Best Local Similarity 5.3%; Score 139; DB 15; Length 26;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 FAPVQNLQWSVSGQTVLTWQAPASD 227
DB 1 FAPVQNLQWSVSGQTVLTWQAPASD 26

RESULT 15

```

```

US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66335
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66335

Query Match
Best Local Similarity 5.2%; Score 138.5; DB 12; Length 2468;
Matches 107; Conservative 47; Mismatches 166; Indels 131; Gaps 23;

QY 103 TIPAGLYDPFEYKVPVNADASESPNFVLDTGTAADIAGTYDYVIINPNPGIIIVGEG 162
DB 319 TDPAG-----NNSTPVTVEAP-----DTTA-----PAPATD-----VQVAPDG 351
QY 163 VS-KGNDYVVEAGTYHFTVQRPQGDAAASVVVTGEGNEFAPVQNLQWSVSGQTVLTW 221
DB 352 SSVTGN--AEPGAT--VGVDTGDDGQPDITVVVPGGSEFEVPLN--PPLTNGETVTVIV 404
QY 222 QAPASDKRTYVLINESFDQTLPN-----GWTMIDADGDCH-----256
DB 405 TDPAGNSSTPVTAEAPDPDPAPQVNASNGSVLSGTAEAGTVITIDGNGNPIGQTSADAN 464
QY 257 -NWLST-----INVI-----NTATH-TGDGAM-----FSKSWTAGS 285
DB 465 GWSFTPGSQLPDGTWVNVVARDAAAGNSPATSITVDGVAPNAPVVPSPNGSELSTAP 524
QY 286 GAKIDLSPDNYLVTPEKTVVPENGKLSYVWSSQVPTNEHYGVFLSTTGNEAANFTIKLE 345
DB 525 GSSVLTLDGNGNPICQTGTADANGNSWFTFSTPLP-----DGTVVNVVARDAAAGNSPPAS 579

```

```
QY 346 ETLSGDKPAPMNLVKSEGVKLPAPYQERTIDLSAYAGQQVYLAFRHENSSTGIERLYLDDV 405
Db 580 VTVDVAVAPATPTVDPSNGT-----TLSGTAEPGSSVTLTDGNGNPIG----- 621
QY 406 AVSGEGSSNDYTYT-----VYRDNVVIAQNLAAITFNOENVAPGQYNYCBEVKYTAGVSPK 461
Db 622 QVTADGSGN-WTFTPTSTPLNGTVW---NATATDPSGNASSPAS-----VTVDVAVAPATPV 673
QY 462 V--CKDVTVEGSNEFAHVQNLIT---GSAVGQ 487
Db 674 VNPSNGTTLSGTAEPGATVTLTDGNGNPIGQ 704
```

Search completed: May 18, 2004, 11:38:27
Job time : 22.8443 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:21:33 ; Search time 87.0274 Seconds

(without alignments)
5623.200 Million cell updates/sec

Title: US-08-353-485-10

Perfect score: 9179

Sequence: 1 MKKLILLIAASLLGLVGLYLAQ.....HYAVMVVVDGKSVKELAVK 1732

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1990s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|------------|--------------------|
| 1 | 9179 | 100.0 | 1732 | 2 AAR96029 | Aar96029 P. gingiv |
| 2 | 9179 | 100.0 | 1732 | 2 AAW24787 | Aaw24787 Ptk anti |
| 3 | 9179 | 100.0 | 1732 | 2 AAW69487 | Aaw69487 Haemagglu |
| 4 | 6151.5 | 67.0 | 1358 | 2 AAR96032 | Aar96032 P. gingiv |
| 5 | 6151.5 | 67.0 | 1358 | 2 AAR96032 | Aar96032 P. gingiv |
| 6 | 4903 | 53.4 | 970 | 2 AAR72458 | Aar72458 Porphyrom |
| 7 | 4673.5 | 50.9 | 1087 | 2 AAR96028 | Aar96028 P. gingiv |
| 8 | 4673.5 | 50.9 | 1087 | 2 AAR96028 | Aar96028 P. gingiv |
| 9 | 4272 | 46.5 | 1706 | 2 AAW24786 | Aaw24786 Ptk anti |
| 10 | 4270 | 46.5 | 1704 | 3 AAY67396 | Aay67396 Arg-gingi |
| 11 | 4270 | 46.5 | 1704 | 4 AAU08938 | Aau08938 P. gingiv |
| 12 | 4264 | 46.5 | 1704 | 2 AAR70188 | Aar70188 Arg-gingi |
| 13 | 4263 | 46.4 | 1704 | 2 AAU34843 | Aau34843 Arg-gingi |
| 14 | 4245 | 46.2 | 1687 | 2 AAW69495 | Aaw69495 Haemagglu |
| 15 | 4233 | 46.1 | 1687 | 2 AAR96033 | Aar96033 P. gingiv |
| 16 | 4217 | 45.9 | 2628 | 2 AAR96030 | Aar96030 P. gingiv |
| 17 | 4217 | 45.9 | 2628 | 2 AAR96030 | Aar96030 P. gingiv |
| 18 | 2618.5 | 28.5 | 506 | 2 AAW69488 | Aaw69488 Haemagglu |
| 19 | 2565 | 27.9 | 509 | 2 AAU83126 | Aau83126 Ptk anti |
| 20 | 2254 | 24.6 | 419 | 4 AAU03574 | Aau03574 P. gingiv |
| 21 | 2034.5 | 22.2 | 456 | 2 AAR96023 | Aar96023 P. gingiv |
| 22 | 2034.5 | 22.2 | 456 | 2 AAR96022 | Aar96022 P. gingiv |
| 23 | 2034.5 | 22.2 | 456 | 2 AAW69491 | Aaw69491 Haemagglu |
| 24 | 2034.5 | 22.2 | 456 | 2 AAW69490 | Aaw69490 Haemagglu |
| 25 | 2025.5 | 22.1 | 439 | 2 AAR96024 | Aar96024 P. gingiv |

| | | | | | |
|----|--------|------|-----|------------|--------------------|
| 26 | 2025.5 | 22.1 | 439 | 2 AAW69492 | Aaw69492 Haemagglu |
| 27 | 2011.5 | 21.9 | 450 | 2 AAR96021 | Aar96021 P. gingiv |
| 28 | 2011.5 | 21.9 | 450 | 2 AAW69489 | Aaw69489 Haemagglu |
| 29 | 1854 | 20.2 | 364 | 2 AAR96031 | Aar96031 P. gingiv |
| 30 | 1854 | 20.2 | 364 | 2 AAW69493 | Aaw69493 Protease |
| 31 | 1776 | 19.3 | 419 | 4 AAU03572 | Aau03572 P. gingiv |
| 32 | 1241 | 13.5 | 231 | 4 AAU03575 | Aau03575 P. gingiv |
| 33 | 1221.5 | 13.3 | 991 | 2 AAR77313 | Aar77313 Porphyrom |
| 34 | 980.5 | 10.7 | 921 | 2 AAY34522 | Aay34522 Porphyrom |
| 35 | 980.5 | 10.7 | 922 | 2 AAY34521 | Aay34521 Porphyrom |
| 36 | 980.5 | 10.7 | 925 | 2 AAY34520 | Aay34520 Porphyrom |
| 37 | 980.5 | 10.7 | 938 | 2 AAY34392 | Aay34392 Porphyrom |
| 38 | 812.5 | 8.9 | 497 | 2 AAR96025 | Aar96025 P. gingiv |
| 39 | 812.5 | 8.9 | 497 | 2 AAW69483 | Aaw69483 Haemagglu |
| 40 | 763 | 8.3 | 148 | 3 AAB14942 | Aab14942 Porphyrom |
| 41 | 749 | 8.2 | 377 | 2 AAY34359 | Aay34359 Porphyrom |
| 42 | 725 | 7.9 | 135 | 6 ABP55081 | Abp55081 Porphyrom |
| 43 | 722 | 7.9 | 312 | 2 AAY34484 | Aay34484 Porphyrom |
| 44 | 715 | 7.8 | 134 | 4 AAB49217 | Aab49217 Peptide u |
| 45 | 698 | 7.6 | 737 | 2 AAR70186 | Aar70186 Arg-gingi |

ALIGNMENTS

RESULT 1

AAR96029
ID AAR96029 standard; protein; 1732 AA.

XX AAR96029;

XX AC

XX 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX P. gingivalis porphyrom.

DE Porphyrom; haemagglutinin; periodontal disease; vaccine; antibody.

XX Porphyromonas gingivalis; strain W12.

XX OS

XX FH Key Location/Qualifiers

FT Region 688..708

FT /note= "Pro-Asn repeat region type 1"

FT Region 887..952

FT /note= "Pro-Asn repeat region type 2"

FT Region 946..967

FT /note= "Pro-Asn repeat region type 1"

FT Region 985..1006

FT /note= "Pro-Asn repeat region type 3"

FT Region 1041..1100

FT /note= "Pro-Asn repeat region type 4"

FT Region 1341..1405

FT /note= "Pro-Asn repeat region type 2"

FT Region 1430..1451

FT /note= "Pro-Asn repeat region type 3"

FT Region 1488..1547

FT /note= "Pro-Asn repeat region type 4"

FT Region 1607..1650

FT /note= "Pro-Asn repeat region type 2"

XX WO9617936-A2.

XX 13-JUN-1996.

XX 11-DEC-1995; 95WO-US016108.

XX 09-DEC-1994; 94US-00353485.

XX (UYEL) UNIV FLORIDA.

XX (UABR-) UAB RES FOUND.

XX Progulskke-Fox A, Tunwasorn S, Legine G, Han N, Lantz M, Patti JM;

XX

DR WPI; 1996-287181/29.
 DR N-PSDB; AAR30653.
 XX
 PT Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 PS Claim 5; Page 76-81; 153pp; English.
 XX
 CC P. gingivalis W12 cysteine protease, porphyrain (AAR96029), was
 CC identified as the product of the prp gene (AAR30653) isolated from P.
 CC gingivalis W12 genomic DNA. The porphyrain shows homology to the
 CC haemagglutinins (see also AAR96036-28 and AAR96030-33) of P. gingivalis
 CC 318. It can be obt. from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows prodn. of a live vaccine. The porphyrain and
 CC haemagglutinins can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 XX Sequence 1732 AA;

Query Match 100.0%; Score 9179; DB 2; Length 1732;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKILLIIAASLLGVLYAQSAKIKLDAPTRTCTNNNSFKQFDASFSFNEVELTKVETK 60
 DB 1 MRKILLIIAASLLGVLYAQSAKIKLDAPTRTCTNNNSFKQFDASFSFNEVELTKVETK 60

QY 61 GGTASVISIPGAFPTGEVGSPEVPAVRKLIAPVPGATPVVRVKSFTEQVYSLNOYGSSEKL 120
 DB 61 GGTASVISIPGAFPTGEVGSPEVPAVRKLIAPVPGATPVVRVKSFTEQVYSLNOYGSSEKL 120

QY 121 MPHOPSKSDDEPKVPVYNAAYARKFGVQBLTQVEMLTGWRGVRIAAALTNPVOYD 180
 DB 121 MPHOPSKSDDEPKVPVYNAAYARKFGVQBLTQVEMLTGWRGVRIAAALTNPVOYD 180

QY 181 VVANQLKVRNNIEVSPQGADEVATQRLYDASFSPFETAYKOLFNRDVYTDHGDLYNT 240
 DB 181 VVANQLKVRNNIEVSPQGADEVATQRLYDASFSPFETAYKOLFNRDVYTDHGDLYNT 240

QY 241 PVRMLVWAGAKFKALPWLTKWAKQGFYLDVHYTDEAEVGTNNASIKAFIHKKYNDGLA 300
 DB 241 PVRMLVWAGAKFKALPWLTKWAKQGFYLDVHYTDEAEVGTNNASIKAFIHKKYNDGLA 300

QY 301 ASAAPVFLALVGDIDVISGEKKTKKYTDLYYSAVDGDEYFEMVTFMSASSPEELTNI 360
 DB 301 ASAAPVFLALVGDIDVISGEKKTKKYTDLYYSAVDGDEYFEMVTFMSASSPEELTNI 360

QY 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSNNSQVGPQTIKYGMQYYNQBHGVTDVVNY 420
 DB 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSNNSQVGPQTIKYGMQYYNQBHGVTDVVNY 420

QY 421 LKAPYTGCSHLNTGVSPANTAHGSETAWADPLLTTSQKALTNKOKYFLAIGNCCITA 480
 DB 421 LKAPYTGCSHLNTGVSPANTAHGSETAWADPLLTTSQKALTNKOKYFLAIGNCCITA 480

QY 481 QFDVYQPCFGEVITRVKEKGAYAYIGSSPNSVWGEDYVWSGVANAVFGVQPTFEGTSMGS 540
 DB 481 QFDVYQPCFGEVITRVKEKGAYAYIGSSPNSVWGEDYVWSGVANAVFGVQPTFEGTSMGS 540

QY 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSMVPYRAMP 600
 DB 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSMVPYRAMP 600

QY 601 KNTNYTLPASLPONCASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
 DB 601 KNTNYTLPASLPONCASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660

QY 661 NYDVVITRSNLYLPVTKIQVGEPSYPQVSNLTATTQGGKVTLLKWEAPSAKKAAGSREVK 720
 DB 661 NYDVVITRSNLYLPVTKIQVGEPSYPQVSNLTATTQGGKVTLLKWEAPSAKKAAGSREVK 720

RESULT 2
 AAR24787
 ID AAR24787 standard; protein; 1732 AA.
 XX

QY 721 RIGDGLFVTTIEPANDVEANEAKVLAADNVGNDTGYQFLLDADHNTFGSVIPATGDLFT 780
 DB 721 RIGDGLFVTTIEPANDVEANEAKVLAADNVGNDTGYQFLLDADHNTFGSVIPATGDLFT 780

QY 781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGGGEVVIPIGGVYDYCIITNPEPASGKWI 840
 DB 781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTGGGEVVIPIGGVYDYCIITNPEPASGKWI 840

QY 841 AGDGNQAPARYDDPTFEAGKKYITFMRAGVGDGTDMEVEDDSDPASVYTYVRDGTIKE 900
 DB 841 AGDGNQAPARYDDPTFEAGKKYITFMRAGVGDGTDMEVEDDSDPASVYTYVRDGTIKE 900

QY 901 GLTATTFEEDGVAAGNHEYCEVVKYTAGVSPKVKDVTVEGSENEFAVQNLTGSSVQKV 960
 DB 901 GLTATTFEEDGVAAGNHEYCEVVKYTAGVSPKVKDVTVEGSENEFAVQNLTGSSVQKV 960

QY 961 TLKWDAPNGTNPNNPNNPNTGTTLSSEFNGIPASWKTTDADGDGHWKFGNAPGLAGY 1020
 DB 961 TLKWDAPNGTNPNNPNNPNTGTTLSSEFNGIPASWKTTDADGDGHWKFGNAPGLAGY 1020

QY 1021 NSNGCVYSSEFGLGGIGVLPDNYLIPTALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080
 DB 1021 NSNGCVYSSEFGLGGIGVLPDNYLIPTALDLPNGGKLTFWVCAQDANYASEHYAVYASS 1080

QY 1081 TGNDASFTNALLLEETITAKGVSPKAI RGRIOGTWROKTVOLPAGTKYVAFRHFQSTDM 1140
 DB 1081 TGNDASFTNALLLEETITAKGVSPKAI RGRIOGTWROKTVOLPAGTKYVAFRHFQSTDM 1140

QY 1141 FYIDLDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGGQGWCLSSGGQLDLTA 1200
 DB 1141 FYIDLDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGGQGWCLSSGGQLDLTA 1200

QY 1201 HGSNNVSSFSWNGMALNPDNYLISKDVTGATKVKYVAYVNDGPPGDHYAVMISKTGTNA 1260
 DB 1201 HGSNNVSSFSWNGMALNPDNYLISKDVTGATKVKYVAYVNDGPPGDHYAVMISKTGTNA 1260

QY 1261 GDFTVFEEEPNGINKGARFGLSTEANGAKPQSVMIERTVDPAGTKYVAFRHYNCSDL 1320
 DB 1261 GDFTVFEEEPNGINKGARFGLSTEANGAKPQSVMIERTVDPAGTKYVAFRHYNCSDL 1320

QY 1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTIKI KEGLTETTFEEDGVATGNHEYCEVVKYT 1380
 DB 1321 NYILLDDIQFTMGGSPTPTDYTYTVYRDGTIKI KEGLTETTFEEDGVATGNHEYCEVVKYT 1380

QY 1381 AGVSPKCVDTVNSVTCFNPQNLTAEOAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
 DB 1381 AGVSPKCVDTVNSVTCFNPQNLTAEOAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

QY 1441 SWKTTIDADGGNNWTTTPPGSSSFAGHNSAICVSSASHINFEQPQNPNDVLTPELSLP 1500
 DB 1441 SWKTTIDADGGNNWTTTPPGSSSFAGHNSAICVSSASHINFEQPQNPNDVLTPELSLP 1500

QY 1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDANFANALLEEVLTAKTVVTAPEAIRGTR 1560
 DB 1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDANFANALLEEVLTAKTVVTAPEAIRGTR 1560

QY 1561 AQTWYQKTVOLPAGTKYVAFRHFQCTDFFWINLDDVITSGNAPSXYTYTYRNTTOIAS 1620
 DB 1561 AQTWYQKTVOLPAGTKYVAFRHFQCTDFFWINLDDVITSGNAPSXYTYTYRNTTOIAS 1620

QY 1621 GVTETTYRDPDLATGFYTYGVKVPYNGESAIEATATLNTSLADVTAQKFTYTLTVVGKTI 1680
 DB 1621 GVTETTYRDPDLATGFYTYGVKVPYNGESAIEATATLNTSLADVTAQKFTYTLTVVGKTI 1680

QY 1681 TVTCQGEAMTYDMNGRRLAAGRNTVVVTAQGGHYAVMVVVDGKSYVEKLVAK 1732
 DB 1681 TVTCQGEAMTYDMNGRRLAAGRNTVVVTAQGGHYAVMVVVDGKSYVEKLVAK 1732

AAW24787;
 17-OCT-2003 (revised)
 25-NOV-1997 (first entry)
 PrtK antigenic protein complex.
 Periodontal disease; cell surface protein; thiol protease; endopeptidase;
 PrtK; PrtK48; PrtK39; PrtK15; PrtK44; haemagglutinin; adhesin; therapy;
 diagnosis; vaccine; antigen.
 Porphyromonas gingivalis; strain W50.
 Key Location/Qualifiers
 Peptide 1..228
 /label= pro-pro_peptide
 Cleavage-site 228..229
 Protein 229..737
 /label= PrtK48
 /note= "48 kDa Lys-specific thiol protease"
 Cleavage-site 737..738
 Protein 738..1156
 /label= PrtK39
 /note= "39 kDa adhesin"
 Cleavage-site 1156..1157
 Protein 1157..1291
 /label= PrtK15
 /note= "15 kDa adhesin"
 Cleavage-site 1291..1292
 Protein 1292..1732
 /label= PrtK44
 /note= "44 kDa adhesin"
 W09176542-Al.
 09-MAY-1997.
 30-OCT-1996; 96WO-AU0000673.
 30-OCT-1995; 95AU-000006275.
 (UYME) UNIV MELBOURNE.
 (VICT-) VICTORIAN DAIRY IND AUTHORITY.
 Reynolds EC, Bhogal PS, Slakeski N;
 WPI: 1997-272112/24.
 N-PSDB; AAT78851.
 New antigenic protein complex from Porphyromonas gingivalis - comprising
 Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
 prevention and treatment of periodontal disease.
 Example 1; Fig 9b; 68pp; English.
 A PrtR-PrtK cell surface protein of Porphyromonas ginivalis (PG)
 comprises a 300 kDa complex composed a 48 kDa lysine-specific thiol
 protease and 39, 15 and 44 kDa adhesins encoded by the prtK gene
 (AAT78851), and a 45 kDa arginine-specific thiol protease and 44, 15, 17
 and 27 kDa adhesins (see AAW24786) encoded by the prtR gene (AAT78850). A
 claimed antigenic complex comprises at least one multimeric protein
 complex of PrtR and PrtK each containing at least one adhesin domain, the
 complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
 proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used
 in a claimed composition to elicit an immune response directed against
 PG, and in a claimed method of reducing the prospect of PG infection
 and/or severity of disease. Antibodies directed against the complex are
 claimed for use in treating PG infection. Unlike whole PG cells or other
 previously prepared antigens based on fimbriae or the capsule, the PrtR-
 PrtK complex or component parts are safe and effective antigens. (Updated
 on 17-OCT-2003 to standardise OS field)
 Sequence 1732 AA;

Query Match 100.0%; Score 9179; DB 2; Length 1732;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRKLLLLIAASLLGLVYAQSAKIKLDAPTRTTCTNNFSKQDFDASFSNEVELTKVETK 60
 DB 1 MRKLLLLIAASLLGLVYAQSAKIKLDAPTRTTCTNNFSKQDFDASFSNEVELTKVETK 60
 QY 61 GGTFAVSIPGAPPTGEVGSPEVPAVRKLIAPVPGATPVVRKVSFTQVYSLNQYSEKL 120
 DB 61 GGTFAVSIPGAPPTGEVGSPEVPAVRKLIAPVPGATPVVRKVSFTQVYSLNQYSEKL 120
 QY 121 MPHQSMKSDDEPEKVPFVYNAAYARKGFVQGLTQVEMGLTMRGVRAALNTINVOYD 180
 DB 121 MPHQSMKSDDEPEKVPFVYNAAYARKGFVQGLTQVEMGLTMRGVRAALNTINVOYD 180
 QY 181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLENRDVYTDHGLYNT 240
 DB 181 VVANQLKVRNNIEIEVSFQGADEVATQRLYDASFSPYFETAYKQLENRDVYTDHGLYNT 240
 QY 241 PVRLVWAGAKFEALKPWLTKWAKGFLVDHYTDEAEVGTNNASIKAFIHKXNDGLA 300
 DB 241 PVRLVWAGAKFEALKPWLTKWAKGFLVDHYTDEAEVGTNNASIKAFIHKXNDGLA 300
 QY 301 ASAAPVFLALVGDTDVISGEKGTTKVTDLYYSADVGDYFPEMYTFRMSASSPELTNI 360
 DB 301 ASAAPVFLALVGDTDVISGEKGTTKVTDLYYSADVGDYFPEMYTFRMSASSPELTNI 360
 QY 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSNMSQVQPTIKYGMQYYNQEHGYTDVYNY 420
 DB 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSNMSQVQPTIKYGMQYYNQEHGYTDVYNY 420
 QY 421 LKAPYTCYSHLNTGVSFANYTAHGETAWADPLTTSOLKALTNDKYFLAIGNCITA 480
 DB 421 LKAPYTCYSHLNTGVSFANYTAHGETAWADPLTTSOLKALTNDKYFLAIGNCITA 480
 QY 481 QFDYVQPCFGEVITRVEKGAAYIGSSPNSWGEDYYSVGANAVFGQPTFEGTSMGS 540
 DB 481 QFDYVQPCFGEVITRVEKGAAYIGSSPNSWGEDYYSVGANAVFGQPTFEGTSMGS 540
 QY 541 YDATELEDSYNTVNSIMMAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSMVPYRAMP 600
 DB 541 YDATELEDSYNTVNSIMMAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDGSMVPYRAMP 600
 QY 601 KNTYTLPASLPQNOASYSIQASAGSYVAISKDGVLVGTGVANASGVATVSMTKQITENG 660
 DB 601 KNTYTLPASLPQNOASYSIQASAGSYVAISKDGVLVGTGVANASGVATVSMTKQITENG 660
 QY 661 NYDVVITRSNVLPIKQIQVGPSPYQVPSNLTATTQGGKVTLKWEAPSAKKAESREVK 720
 DB 661 NYDVVITRSNVLPIKQIQVGPSPYQVPSNLTATTQGGKVTLKWEAPSAKKAESREVK 720
 QY 721 RIGDGLFVTIEPANDVRANEAKVLAADNVMDGNTGYQFLLDADHNTFGSVIPATGPLFT 780
 DB 721 RIGDGLFVTIEPANDVRANEAKVLAADNVMDGNTGYQFLLDADHNTFGSVIPATGPLFT 780
 QY 781 GTASNLNLSANFEYLAVPANADPVVTTQMIIVTQGEVVI PGGVYDYCITNPEPAGKMWI 840
 DB 781 GTASNLNLSANFEYLAVPANADPVVTTQMIIVTQGEVVI PGGVYDYCITNPEPAGKMWI 840
 QY 841 AGDGGNQPARYDDFTFEAGKGYTFMRBAGMGDGTDMVEDDPSASYTYTVYRDGTCKI 900
 DB 841 AGDGGNQPARYDDFTFEAGKGYTFMRBAGMGDGTDMVEDDPSASYTYTVYRDGTCKI 900
 QY 901 GLTATTFEEDGVAAGNHBYCVVEVKYTAGVSPKCVKDVTVGSNEFAPQNLTGSSVGQKV 960
 DB 901 GLTATTFEEDGVAAGNHBYCVVEVKYTAGVSPKCVKDVTVGSNEFAPQNLTGSSVGQKV 960
 QY 961 TLKWDAPNGTNPNNPNNPNTGTTLSFENGIPASWKTIDADGCHGWKGNAPGIAGY 1020
 DB 961 TLKWDAPNGTNPNNPNNPNTGTTLSFENGIPASWKTIDADGCHGWKGNAPGIAGY 1020

Sequence 1732 AA;

QY 1021 NSNGCVYSESGLGIGVLTDPNVLITPALDLPNGKLTFFWCAQDANYASEHVAAYASS 1080
 Db |||||
 1021 NSNGCVYSESGLGIGVLTDPNVLITPALDLPNGKLTFFWCAQDANYASEHVAAYASS 1080
 QY 1081 TGNDASNTNALLEETIAKGVRSKPAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDTM 1140
 Db |||||
 1081 TGNDASNTNALLEETIAKGVRSKPAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDTM 1140
 QY 1141 FYIDLDEVEIKANGKRAPDTTFESSTHGEPAEWTTIDADGGGWLCLSSGOLDWLTA 1200
 Db |||||
 1141 FYIDLDEVEIKANGKRAPDTTFESSTHGEPAEWTTIDADGGGWLCLSSGOLDWLTA 1200
 QY 1201 HGSNVVSFSWNGMALNPDNYLISKVDTGATKVKYIYAVNDGFGPDHYAVMISKTGNA 1260
 Db |||||
 1201 HGSNVVSFSWNGMALNPDNYLISKVDTGATKVKYIYAVNDGFGPDHYAVMISKTGNA 1260
 QY 1261 GDFTVVFEETPNKNGARFGLSTEANGAKPQSVWIERVTDLDPAGTKYVAFRHYNCSDL 1320
 Db |||||
 1261 GDFTVVFEETPNKNGARFGLSTEANGAKPQSVWIERVTDLDPAGTKYVAFRHYNCSDL 1320
 QY 1321 NYILDDIQTWGGSPPTDVTYTVYRDGTIKKEGLTETTFEEDGVAAGNHEYCVVEKYT 1380
 Db |||||
 1321 NYILDDIQTWGGSPPTDVTYTVYRDGTIKKEGLTETTFEEDGVAAGNHEYCVVEKYT 1380
 QY 1381 AGVSPKCKVDVTNSTQPNVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDPENGIPA 1440
 Db |||||
 1381 AGVSPKCKVDVTNSTQPNVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDPENGIPA 1440
 QY 1441 SWKTIADADGNNWTTTPPGSSPAGHNSAICVSSASHINFEQPNPDNLVLTPELSLP 1500
 Db |||||
 1441 SWKTIADADGNNWTTTPPGSSPAGHNSAICVSSASHINFEQPNPDNLVLTPELSLP 1500
 QY 1501 GGGTLTFWVCAQDANYASEHVAAYVASTGNDASNPANALLEVLTAKTVVTAPPAIRGTR 1560
 Db |||||
 1501 GGGTLTFWVCAQDANYASEHVAAYVASTGNDASNPANALLEVLTAKTVVTAPPAIRGTR 1560
 QY 1561 AQTWYQKTVDLPAGTKYVAFRHFCTDFFWINLDDVVITSGNAPSYYTIIYRNNQTIAS 1620
 Db |||||
 1561 AQTWYQKTVDLPAGTKYVAFRHFCTDFFWINLDDVVITSGNAPSYYTIIYRNNQTIAS 1620
 QY 1621 GVTETTRDPLATGFTYTGKVVYPNGESAIEATLNITSLADVTAKPYTLTVGKTI 1680
 Db |||||
 1621 GVTETTRDPLATGFTYTGKVVYPNGESAIEATLNITSLADVTAKPYTLTVGKTI 1680
 QY 1681 TVTCQGEAMIYDMGRRLAAGRTVVTYTAQGHVAVMVVDGKSVVEKLVK 1732
 Db |||||
 1681 TVTCQGEAMIYDMGRRLAAGRTVVTYTAQGHVAVMVVDGKSVVEKLVK 1732

RESULT 3

AAW69487
 ID AAW69487 standard; protein; 1732 AA.

AC AAW69487;

DT 22-DEC-1998 (first entry)

DE Haemagglutinin protein prtP.

KW Haemagglutinin protein; periodontal disease; vaccine; prtP.

OS Porphyromonas gingivalis.

FN US5824791-A.

PD 20-OCT-1998.

FF 11-DEC-1995; 95US-00570311.

PR 08-SEP-1988; 88US-00241640.

PR 25-JAN-1991; 91US-00647119.

PR 09-DEC-1994; 94US-00353485.

XX

PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 PI Patti JM, Han N, Lantz M, Tumwasorn S, Proguliske-Fox A, Lepine G;
 XX WPI; 1998-582627/49.
 DR N-PSDB; AAV58874.
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 protease poly:peptide(s)).
 XX Claim 1; Col 69-84; 101pp; English.
 XX This sequence is encoded by a Porphyromonas gingivalis gene of the
 invention. This sequence represents the prtP haemagglutinin protein. The
 polypeptides are used to produce antibodies to organisms associated with
 periodontal disease. The antibodies are also used in purification and
 identification procedures. The genes and polypeptides are used as
 vaccines against periodontal disease
 XX Sequence 1732 AA;

Query Match 100.0%; Score 9179; DB 2; Length 1732;
 Best Local Similarity 100.0%; Pred.No. 0;
 Matches 1732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLLILLIAASLLGVGLYQAQSAKIKLDAPTRTCTTNSFKQPDASFSFNEVELTKVETK 60
 Db |||||
 1 MRKLLILLIAASLLGVGLYQAQSAKIKLDAPTRTCTTNSFKQPDASFSFNEVELTKVETK 60
 QY 61 GGTFFASVSIIGAPFTGSEVGPVPAVRKLIAPVGAATPVVRKSFTEQVSLNYSYSEKL 120
 Db |||||
 61 GGTFFASVSIIGAPFTGSEVGPVPAVRKLIAPVGAATPVVRKSFTEQVSLNYSYSEKL 120
 QY 121 MPHQPSMSKSDDBPKVPFYVNAAYARKFVQBELTQVEMLTMRGVRIIAALTINPVQYD 180
 Db |||||
 121 MPHQPSMSKSDDBPKVPFYVNAAYARKFVQBELTQVEMLTMRGVRIIAALTINPVQYD 180
 QY 181 VVANQLKVRNNIEVLSFGQADEVATQRLYDASPSYFETAYKQLFNRDVTYTHGDLINT 240
 Db |||||
 181 VVANQLKVRNNIEVLSFGQADEVATQRLYDASPSYFETAYKQLFNRDVTYTHGDLINT 240
 QY 241 PVRLMVVAGAKFKEALKPWLTKWAKQGFYLDVHTDEAEVGTTNASTKAFTHKKYNDGLA 300
 Db |||||
 241 PVRLMVVAGAKFKEALKPWLTKWAKQGFYLDVHTDEAEVGTTNASTKAFTHKKYNDGLA 300
 QY 301 ASAPFVFLALVGDVTVISGEGKKTKKVTDLYSAVDGYFPEPMYTFRMSASSPEELTNI 360
 Db |||||
 301 ASAPFVFLALVGDVTVISGEGKKTKKVTDLYSAVDGYFPEPMYTFRMSASSPEELTNI 360
 QY 361 IDKVLMYEKATMPDKSVLEKVLIIAGADYSWNSQVQPTIKYGMYYNOEHGYTDVYNY 420
 Db |||||
 361 IDKVLMYEKATMPDKSVLEKVLIIAGADYSWNSQVQPTIKYGMYYNOEHGYTDVYNY 420
 QY 421 LKAPYTCYSHLNTGVSFANYTAHGSSETAWADPLLTTSQLKALTNDKDYFLAIGNCCITA 480
 Db |||||
 421 LKAPYTCYSHLNTGVSFANYTAHGSSETAWADPLLTTSQLKALTNDKDYFLAIGNCCITA 480
 QY 481 QFDYVQCFGEVITRVKEKAYAYIGSSPNYSYWGEDIYNSVGANAVFGVQPTFEGTSMGS 540
 Db |||||
 481 QFDYVQCFGEVITRVKEKAYAYIGSSPNYSYWGEDIYNSVGANAVFGVQPTFEGTSMGS 540
 QY 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITTHIGAHYYWEAYHVLGSGVMPYRAMP 600
 Db |||||
 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITTHIGAHYYWEAYHVLGSGVMPYRAMP 600
 QY 601 KTNITYTLPASLPQNASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
 Db |||||
 601 KTNITYTLPASLPQNASYSIQASAGSYVAISKDGVLYGTGVANASGVATVSMTKQITENG 660
 QY 661 NYDWTITRSNYLPIVKIQIQTGEPSPQVPSNLTATTQGGKQVTLKWBAPSAKKAGSREVK 720
 Db |||||
 661 NYDWTITRSNYLPIVKIQIQTGEPSPQVPSNLTATTQGGKQVTLKWBAPSAKKAGSREVK 720


```
QY 721 RIGDGLFVTEIPANDVRANEAKVLLAADNVGNTQYQFLLDADHNTFGSVLPATGPLET 780
DB 721 RIGDGLFVTEIPANDVRANEAKVLLAADNVGNTQYQFLLDADHNTFGSVLPATGPLET 780
QY 781 GTASSNLYSANFEYLVPANADPVTTTQNIIVTQGGSEWIPGGVYDYCIITNPEPASGKMI 840
DB 781 GTASSNLYSANFEYLVPANADPVTTTQNIIVTQGGSEWIPGGVYDYCIITNPEPASGKMI 840
QY 841 AGDGGNQPARYDDTFEAGKKYTFTRRAGMGDGTDMVEDDSPASYTYTVVRDGTGKKE 900
DB 841 AGDGGNQPARYDDTFEAGKKYTFTRRAGMGDGTDMVEDDSPASYTYTVVRDGTGKKE 900
QY 901 GLTATTFEEDGVAAGNHEVCVEVKYTAGVSPKCKDVTVEGSENEFAPVQNLGTSSVGQKV 960
DB 901 GLTATTFEEDGVAAGNHEVCVEVKYTAGVSPKCKDVTVEGSENEFAPVQNLGTSSVGQKV 960
QY 961 TLKWDAPNCTPNPNPNPNPGTTLSESPFENGIPASWKTIDADGDGHGKPGNAPGIAGY 1020
DB 961 TLKWDAPNCTPNPNPNPNPGTTLSESPFENGIPASWKTIDADGDGHGKPGNAPGIAGY 1020
QY 1021 NSNGCVYSEFGLGGIGLVLPDNYLTPALDLPNGKLTFWVCAQDANYASEHYAVYASS 1080
DB 1021 NSNGCVYSEFGLGGIGLVLPDNYLTPALDLPNGKLTFWVCAQDANYASEHYAVYASS 1080
QY 1081 TGNDAENFTNALLEETITAKGVRSPKAIIRGRIQGTWRQKTVDLPAGTKYVAPRHFPOSTDM 1140
DB 1081 TGNDAENFTNALLEETITAKGVRSPKAIIRGRIQGTWRQKTVDLPAGTKYVAPRHFPOSTDM 1140
QY 1141 FYIDLDEIVEIKANGKREADTFETPESSTHGEAPAEWTTIDADGGQGLCLSSQGLDLTA 1200
DB 1141 FYIDLDEIVEIKANGKREADTFETPESSTHGEAPAEWTTIDADGGQGLCLSSQGLDLTA 1200
QY 1201 HGSNNVVSFSWNGMALNPNDYILSKDVTGATKVKYVYAVNDGFPDGHYAVMLSKGTNA 1260
DB 1201 HGSNNVVSFSWNGMALNPNDYILSKDVTGATKVKYVYAVNDGFPDGHYAVMLSKGTNA 1260
QY 1261 GDTVVFEETPNKINGGARFGLSTANGAKPOSVMWERTVDLPAGTKYVAPRHYNCSDL 1320
DB 1261 GDTVVFEETPNKINGGARFGLSTANGAKPOSVMWERTVDLPAGTKYVAPRHYNCSDL 1320
QY 1321 NYILLDDIQTWGGSPPTDYTVTVVRDGTGKKEGLTETTFEEDGVATGNHGYCVRVKYT 1380
DB 1321 NYILLDDIQTWGGSPPTDYTVTVVRDGTGKKEGLTETTFEEDGVATGNHGYCVRVKYT 1380
QY 1381 AGVSPKKCVDDVTNSTQFNPVQNLTAEQAPNSMDAILKNAPASKAEVLNEDFENGIEA 1440
DB 1381 AGVSPKKCVDDVTNSTQFNPVQNLTAEQAPNSMDAILKNAPASKAEVLNEDFENGIEA 1440
QY 1441 SWKTTIDADGDGNNTWTTTPPGSSFAGHNSAICVSSASHINFEQPONPNYLVTPELSLP 1500
DB 1441 SWKTTIDADGDGNNTWTTTPPGSSFAGHNSAICVSSASHINFEQPONPNYLVTPELSLP 1500
QY 1501 GGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANALLEEVLTAKTVVTVAPEAIRGTR 1560
DB 1501 GGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANALLEEVLTAKTVVTVAPEAIRGTR 1560
QY 1561 AOGTWQKTQVLPAQTKYVAFRFGCTDFWNLDDWITSGNAPSYYTIIYRNNTQIAS 1620
DB 1561 AOGTWQKTQVLPAQTKYVAFRFGCTDFWNLDDWITSGNAPSYYTIIYRNNTQIAS 1620
QY 1621 GVTETTYRDPDLATGYTYGVKVVYPNGESAIEATLNTITSLADVTAQKPYTLTVVGKTI 1680
DB 1621 GVTETTYRDPDLATGYTYGVKVVYPNGESAIEATLNTITSLADVTAQKPYTLTVVGKTI 1680
QY 1681 TVTCQGEAMTYDMNGRRLAAGRNTVYVTAQGGHYAVMVVVDGKSYVEKLAVK 1732
DB 1681 TVTCQGEAMTYDMNGRRLAAGRNTVYVTAQGGHYAVMVVVDGKSYVEKLAVK 1732
```

RESULT 4

AAR96032

ID AAR96032 standard; protein; 1358 AA.

```
XX AAR96032;
AC 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX P. gingivalis hagD haemagglutinin.
XX Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
XX Porphyromonas gingivalis; strain FDC381.
XX WO9617936-A2.
XX 13-JUN-1996.
XX 11-DEC-1995; 95WO-US016108.
XX 09-DEC-1994; 94US-00353485.
XX (UYEL) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX WPI: 1996-287181/29.
XX N-PSDB; AAT30655.
XX Porphyromonas gingivalis genes and proteins - used in the detection and
XX vaccination against periodontal disease.
XX Claim 5; Page 125-129; 153pp; English.
XX P. gingivalis 381 haemagglutinin hagD (AAR96032) was identified as the
XX product of the second open reading frame of the hagD gene (AAR30655)
XX derived from P. gingivalis 318 genomic DNA. A first open reading frame
XX coded for hagD protease (see also AAR96031). The protease and
XX haemagglutinin can be obtd. from transformed host cells and used in
XX vaccines to protect humans or animals against periodontal disease.
XX Expression in Salmonella cells allows prodn. of live vaccines. The
XX haemagglutinin and protease can also be used to detect the presence of
XX anti-P. gingivalis antibodies and to raise monoclonal antibodies for
XX diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX Sequence 1358 AA;
Query Match 67.0%; Score 6151.5; DB 2; Length 1358;
Best Local Similarity 85.0%; Pred. NO. 0;
Matches 1170; Conservative 60; Mismatches 118; Indels 29; Gaps 10;
QY 366 MYEKATMPDKSYLEKVLITAGADYSWNSQVQPTIKYGMYYNOEHGYTDVNYLYKAPY 425
DB 1 MYEKATMPDKSYLEKALLIAGADSYWNPKIQQOTIKIYQVYINQDHGYTDVTSYPAPY 60
QY 426 TGCYSHLNTGVSFANYTAHGETAWADPLLTTSQKALTNKDKYFLAIGNCCITTAQFDYV 485
DB 61 TGCYSHLNTGVFANYTAHGETSWADPSLTATQVKTATQKFLAIGNCCVTAQFDYP 120
QY 486 QPCFEVITRVKEKAYAYIGSSPNSYNGEDYVWSVGANAVFGVQPTFEGTSMGSDATF 545
DB 121 QPCFGEVITRVKEKAYAYIGSSPNSYNGEDYVWSVGANAVFGVQPTFEGTSMGSDATF 180
QY 546 LEDSYNTVNSIMWAGNLAATHAGNIGNITHICAHYWEAHYVHLGDSVMPYRAMPKNTNY 605
DB 181 LEDSYNTVNSIMWAGNLAATHAGNIGNITHICAHYWEAHYVHLGDSVMPYRAMPKNTNY 240
QY 606 TLPSLIPQOQASYISQASAGSYVAISKDGLVGTGVANASGVATVSWTKQITENGNDYDV 665
DB 241 TLPSLIPQOQASYISQASAGSYVAISKDGLVGTGVANASGVATVSWTKQITENGNDYDV 300
QY 666 ITRSNYLPVTKIOVGEPSYPQVSNLTATTQGGKVTLKWEAPSAKKAEGSREVKRIGDG 725
DB 301 ITRSNYLPVTKIOVGEPSYPQVSNLTATTQGGKVTLKWEAPSAKKAEGSREVKRIGDG 360
```

```

Qy 726 LFTVIEPANDVRANEAKVVLAAADNVWGNTGVCQFLDADHNTFGSVPATGFLFTGTASS 785
Db 361 LFTVIEPANDVRANEAKVVLAAADNVWGNTGVCQFLDADHNTFGSVPATGFLFTGTASS 420
Qy 786 NLYSANFEYLVPANADPVVTTQNIIVTQGQEWIIPGGVYDYCIITNPBPASGKMWIAGDGG 845
Db 421 NLYSANFEYLVPANADPVVTTQNIIVTQGQEWIIPGGVYDYCIITNPBPASGKMWIAGDGG 480
Qy 846 NOPARYDDTFEAGKKTFTWRAGMGDCTMEVEDDPSASYTYTVYRDGKIKEGLTAT 905
Db 481 NOPARYDDTFEAGKKTFTWRAGMGDCTMEVEDDPSASYTYTVYRDGKIKEGLTAT 540
Qy 906 TFEEDGVAAGNHEVCVEVKYTAGVSPKVKCDVTVEGSNEFAPVQNLTGSSVGQKVTLKWD 965
Db 541 TFEEDGVAAGNHEVCVEVKYTAGVSPKVKCDVTVEGSNEFAPVQNLTGSAVGQKVTLKWD 600
Qy 966 APNGTNPENPNPNTGTTLSSEFENGIPASWKTIADGDGHWKPGNAPGIAGYNSNGC 1025
Db 601 APNGTNPENPNPNTGTTLSSEFENGIPASWKTIADGDGHWKPGNAPGIAGYNSNGC 659
Qy 1026 VYSESGLGIGLVLPDNYLTLPALDLPNGGKLTFWCAQDANYASHYAYASSTGND 1085
Db 660 VYSESGLGIGLVLPDNYLTLPALDLPNGGKLTFWCAQDANYASHYAYASSTGND 719
Qy 1086 SNFTNALLEETITAKGVRSPKAIIRIGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 1145
Db 720 SNFTNALLEETITAKGVRSPKAIIRIGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDL 779
Qy 1146 DEVEIKANGRADTFEFTESSTHGEAPAEWTTIADGDGQWGLCSGQDLWLTAGHGSN 1205
Db 780 DEVEIKANGRADTFEFTESSTHGEAPAEWTTIADGDGQWGLCSGQDLWLTAGHGSN 839
Qy 1206 VVSFSFNGMALNPNDYLSKDVATGATKVKYVAVNDGFPDGHVAVMSKGTNAGDFTV 1265
Db 840 VVSFSFNGMALNPNDYLSKDVATGATKVKYVAVNDGFPDGHVAVMSKGTNAGDFTV 899
Qy 1266 VFEETPNGINKGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNILL 1325
Db 900 VFEETPNGINKGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNILL 959
Qy 1326 DDIOFTWGSSTPDYTYTVYRDGKIKEGLTETTFEEDGVATGNHVCVEVKYTAGVSP 1385
Db 960 DDIOFTWGSSTPDYTYTVYRDGKIKEGLTETTFEEDGVATGNHVCVEVKYTAGVSP 1019
Qy 1386 KKCVDVTVNSQFNPVONLTAEQAPNSMDALIKWNAPASRAEVLNEDFE-NGIPASWKT 1444
Db 1020 KVCNVNLTINPTQFNPVKNLKAQ--PDGDGVVLKWEAPSGKRGELLNEDFEGDALPTGTA 1077
Qy 1445 IDADGDGNW--TTTPPPGG-----SSPAGHNSAICVSSASHINFEQPNPDNLYLTPEL 1497
Db 1078 LDADGDGNWMDITLNEETRGRHVLSPLRASNVASISYSLLOQGEYLPLTENNELITPKV 1137
Qy 1498 SLPGGGTITFWCAQD--ANVASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPAEI 1556
Db 1138 E--GAKKITYKVGSGPLQWMSHDHYALCISKSGTAAADF-----EVIFEETMTYTOGGA 1189
Qy 1557 RGTRAQGTWYQKTQVLPAQTKYVAFRHFQDCTDFWNLDDVWIT-SGNAPSYTYTVIRNN 1615
Db 1190 NLTREK-----DLPAQTKYVAFRHYNCIDVLGIMDDVWITGEGEGSYTYTVYRDG 1241
Qy 1616 TQIASGVTTETTRPDLATGPTTYGVKVPYNGSSAIEATFLNITSADVTAQKPYTLTV 1675
Db 1242 TKIQEGLTETTYTRDAGMSAQHSEYCEVYKAAVSPKVCVDYIPDGVAADVTAQKPYTLTV 1301
Qy 1676 VGTITVTCQGEAMTYDMNGRRLAAGRTVVYTAQGGHYAVMWVVDGKSYVEKLAVK 1732
Db 1302 VGTITVTCQGEAMTYDMNGRRLAAGRTVVYTAQGGHYAVMWVVDGKSYVEKLAIK 1358

```

RESULT 5

AAW69494

ID AAW69494 standard; protein; 1358 AA.

```

XX AAW69494;
XX AC
XX 22-DEC-1998 (first entry)
XX DE
XX Haemagglutinin protein hagD.
XX KW Haemagglutinin protein; periodontal disease; vaccine; hagD.
XX OS Porphyromonas gingivalis.
XX PN US5824791-A.
XX PD 20-OCT-1998.
XX PF 11-DEC-1995; 95US-00570311.
XX PR 08-SEP-1988; 88US-00241640.
XX PR 25-JAN-1991; 91US-00647119.
XX PR 09-DEC-1994; 94US-00353485.
XX PA (UYFL) UNIV FLORIDA.
XX PA (UABR-) UAB RES FOUND.
XX PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulskie-Fox A, Lepine G;
XX WPI: 1998-582627/49.
XX N-PSDB; AAV58880.
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly:peptide(s)).
XX Claim 1; Col 145-158; 101pp; English.
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX invention. This sequence represents the hagD haemagglutinin protein. The
XX polypeptides are used to produce antibodies to organisms associated with
XX periodontal disease. The antibodies are also used in purification and
XX identification procedures. The genes and polypeptides are used as
XX vaccines against periodontal disease
XX SQ Sequence 1358 AA;
Query Match 67.0%; Score 6151.5; DB 2; Length 1358;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 1170; Conservative 60; Mismatches 118; Indels 29; Gaps 10;
Qy 366 MYEKATWPDKSYLEKULLIAGADYSNWSQVQPTIKYGMQYYNQEHGYTDVINYLKAPY 425
Db 1 MYEKATWPDKSYLEKALLIAGADSYWNPKTGQOTIKYAVQYYNQDHGYTDVYSYPKPY 60
Qy 426 TGCYSHLNTGVSFANYTAHGETAWADPLLTTSQKALTNDKVFIAIGNCCITAOFDYV 485
Db 61 TGCYSHLNTGVSFANYTAHGETSWADPLTATQVKTALNDKVFIAIGNCCVTAQFDYP 120
Qy 486 QPCFGEVITRVKEKGAAYIYIGSSPNSYWGEDYVWSGANAVFGVQPTFECTSMGSDATF 545
Db 121 QPCFGEVITRVKEKGAAYIYIGSSPNSYWGEDYVWSGANAVFGVQPTFECTSMGSDATF 180
Qy 546 LEUSYNTVNSIMWAGNLAATHAGNIGNITHIGHYWEAYHVLGDGSVMPYRAMPKNTY 605
Db 181 LEUSYNTVNSIMWAGNLAATHAGNIGNITHIGHYWEAYHVLGDGSVMPYRAMPKNTY 240
Qy 606 TLPASLPQNASYSIOASAGSYVAISKDGLYCTGVANAGSVATVSMTKOITENGNDYDV 665
Db 241 TLPASLPQNASYSIOASAGSYVAISKDGLYCTGVANAGSVATVSMTKOITENGNDYDV 300
Qy 666 ITRSNLVPVTKIQVGEPSYPQVSNLTATTOCKVTLKWEAPSAKKAESRVEKRGIGD 725
Db 301 ITRSNLVPVTKIQAGEPSYPQVSNLTATTOCKVTLKWDAPSAKKAESRVEKRGIGD 360
Qy 726 LFTVIEPANDVRANEAKVVLAAADNVWGNTGVCQFLDADHNTFGSVPATGFLFTGTASS 785

```

Db 361 LFTIEPANDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLEFTCTASS 420
QY 786 NLYSANFEYLVPANADPVVTQNIIVTGOGEVVPVGGVYDICTNPEPAGKMWIAGDGG 845
Db 421 NLYSANFEYLVPANADPVVTQNIIVTGOGEVVPVGGVYDICTNPEPAGKMWIAGDGG 480
QY 846 NOPARYDDTFFAAGKYYTFTMRAGMGDGTDEVEDDSPA SYTYTVYVRDGTIKKEGLTAT 905
Db 481 NOPARYDDTFFAAGKYYTFTMRAGMGDGTDEVEDDSPA SYTYTVYVRDGTIKKEGLTAT 540
QY 906 TFEEDGVAAGNHEYCEVEKYTAGVSPKVKDVTVEGSENFAPVQNLTGSSVGGQVTLKWD 965
Db 541 TFEEDGVAAGNHEYCEVEKYTAGVSPKVKDVTVEGSENFAPVQNLTGSAVGQVTLKWD 600
QY 966 APNGTNPENPNPNPGTTLSESFENGIPASWKTIDADGGHGWKPGNAPGIAGYNSGC 1025
Db 601 APNGTNPENPNPNPGTTLSESFENGIPASWKTIDADGGHGWKPGNAPGIAGYNSGC 659
QY 1026 VYSESGGLGIGVLTDPNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGND 1085
Db 660 VYSESGGLGIGVLTDPNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGND 719
QY 1086 SNFTNALLBETITAKGVRSKAIIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDTDFYIDL 1145
Db 720 SNFTNALLBETITAKGVRSKAIIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDTDFYIDL 779
QY 1146 DEVEIKANGKRA DFTETFESSHTGEAPAEWTTIDADGGGWLCLSSGOLDMLTAHGGSN 1205
Db 780 DEVEIKANGKRA DFTETFESSHTGEAPAEWTTIDADGGGWLCLSSGOLDMLTAHGGSN 839
QY 1206 WVSSFNGMGNALPNLISKDVTGATKVKYVAVNDGFGPDHAYVMSKTGTNAGDFTV 1265
Db 840 WVSSFNGMGNALPNLISKDVTGATKVKYVAVNDGFGPDHAYVMSKTGTNAGDFTV 899
QY 1266 VFEETENGINKGARGLSTEANGAPQSVWIBRTVDLPAGTKYVAFRHYNCSDMLNYILL 1325
Db 900 VFEETENGINKGARGLSTEANGAPQSVWIBRTVDLPAGTKYVAFRHYNCSDMLNYILL 959
QY 1326 DDIOFTMGSSPTDYTYTVYRDGTIKKEGLTTFEEDGVATGNHEYCEVEKYTAGVSP 1385
Db 960 DDIOFTMGSSPTDYTYTVYRDGTIKKEGLTTFEEDGVATGNHEYCEVEKYTAGVSP 1019
QY 1386 KRCVDVTNOSTQPNVQNLTAEOAPNSMDAILKWNAPASKRAEVLNEDFE-NGIPASWKT 1444
Db 1020 KVCNVNTINPTQPNVKNLKAQ--PDGGDVLKWEAPSGKRGLNEDPEGDAIPGTWTA 1077
QY 1445 IDADGNNW--TTTPPGG-----SSPAGHNSAICVSSASHINFEQPONPNLYTPEL 1497
Db 1078 LDADGGNNWDTLNEFTGERHVLSPLRASNVASISLLOQGEYLPNTNNFLITPKV 1137
QY 1498 SLPGGGTLTFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEVLTAKTVVTAPAI 1556
Db 1138 E--GAKIYKYVSGPLQWSDHSHALCLSKSTAADP-----EVIETNTYTOGGA 1189
QY 1557 RGTAGQTYQKTVQLPAGTKYVAFRHFQCTDFWNLDDWIT--SGNAPSYYTYTYRNN 1615
Db 1190 NLTREK-----DLPAGTKYVAFRHYNCTDVLGIMIDWVITGREGESYTYTVYRDG 1241
QY 1616 TQIASQVTTTTRDPLATGFTYGVKVPNGESAIEATLNTSLADTAQKPYTLTV 1675
Db 1242 TKIQEGLTETTRDAGMSAQSHYCEVEKYAGVSPKVCVDYIPDGVADTAQKPYTLTV 1301
QY 1676 VGKTIIVTCQGEAMIVDMNGRRLAAGNTVVTYTAQGHYAVMVVDPGKSVYEKLVK 1732
Db 1302 VGKTIIVTCQGEAMIVDMNGRRLAAGNTVVTYTAQGHYAVMVVDPGKSVYEKLVK 1358

RESULT 6

AAR72458

ID AAR72458 standard; protein; 970 AA.

XX AAR72458;

XX

DT 19-DBC-1995 (first entry)
XX Porphyromonas gingivalis lysine-gingipain protein complex.
DE lysine-gingipain protein complex; amidolytic; proteolytic;
XX lysine specific proteinase; modulator identification; periodontitis;
KW therapy monitoring.
KW Porphyromonas gingivalis.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..228
FT /label= sig_peptide
FT Peptide 229..970
FT /label= mat_peptide
XX
XX W09511298-A1.
PN 27-APR-1995.
XX
PD 21-OCT-1994; 94WO-US012094.
XX
PF 21-OCT-1993; 93US-00141324.
XX
PR (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX Travis J, Potempa JS, Barr P, Pavloff N, Pike RN;
XX WPI; 1995-170220/22.
DR N-PSDB; AAQ88141.
XX
XX Lys-gingipain complex prepn with amidolytic and proteolytic specificity -
PT for cleavage of an amide bond with Lysine contributing the carboxyl gp.
XX
XX Claim 2; Page 53-57; 75pp; English.
XX
XX AAQ88141 encodes AAR72458 the Porphyromonas gingivalis lysine-gingipain
CC protein complex (IGPC). The IGPC has amidolytic and proteolytic
CC specificity for an amide bond, where lysine contributes the carboxyl gp.,
CC i.e. a lysine specific protease. IGPC can be used to identify agents that
CC modulate the effect of IGPC on animals, and also for monitoring the
CC exposure of an animal to IGPC. Such a method can be used to monitor the
CC progress of a therapy designed to lessen the symptoms of periodontitis
XX
XX Sequence 970 AA;
SQ
Query Match 53.4%; Score 4903; DB 2; Length 970;
Best Local Similarity 95.6%; Pred. No. 1.7e-316;
Matches 927; Conservative 21; Mismatches 22; Indels 0; Gaps 0;
QY 1 MRKLLLLIAASLLGVLGVAQSAKIKLDAPTTCTNNSTKOPDASFSFNEVELTKVETK 60
Db 1 MRKLLLLIAASLLGVLGVAQNAKIKLDAPTTCTNNSTKOPDASFSFNEVELTKVETK 60
QY 61 GGTFASVSPGAPFTGEVSGPEVPAVKLIAPVPGATPVVRVKSFTQVYSLNQYSEKL 120
Db 61 GGTFASVSPGAPFTGEVSGPEVPAVKLIAPVPGATPVVRVKSFTQVYSLNQYSEKL 120
QY 121 MPHQSMSKDDPEKVPFVYNAAAARVAKGFVGOELTQVEMGLTMRGVRIAAALTINPVQYD 180
Db 121 MPHQSMSKDDPEKLPFPAYNAAAARVAKGFVGOELTQVEMGLTMRGVRIAAALTINPVQYD 180
QY 181 VVANOLKVRNNTEIEVSFGQDADEVATORLYDASFSYPFETAYKQLFNRDVTYDHGDLNT 240
Db 181 VVANOLKVRNNTEIEVSFGQDADEVATORLYDASFSYPFETAYKQLFNRDVTYDHGDLNT 240
QY 241 PYRMLVAGAKFEALKPWLTKAQGFYLDVHYTDEAEVGTNNASIKAFIHKKNDGLA 300
Db 241 PYRMLVAGAKFEALKPWLTKAQGFYLDVHYTDEAEVGTNNASIKAFIHKKNDGLA 300
QY 301 ASAAPVFLALVGDTDIVISGEKGGTKKVTDLYYSAVDGDFPFEMTYTFRMSASSPELTNI 360
Db 301 ATAAPVFLALVGDTDIVISGEKGGTKKVTDLYYSAVDGDFPFEMTYTFRMSASSPELTNI 360

QY 361 IDKVLMEKATPDKSYLEKVLIIAGADYSWNSQVGPQTIKYGMQYNNQEHGYTDVYNY 420
 Db IDKVLMEKATPDKSYLEKVLIIAGADYSWNSQVGPQTIKYGMQYNNQEHGYTDVYNY 420
 QY 421 LKAPYTGCSYHLNTGVSFANYTAHGETAWADPLLTSQLKALTNDKCYFLAIGNCCTFA 480
 Db PRAPYTGCSYHLNTGVSFANYTAHGETAWADPLLTSQLKALTNDKCYFLAIGNCCTFA 480
 QY 481 QFYDYPQCFGEVITRVEKGAAYVIGSSPNSYWGEDYVMSVGNAGVFGVQPTPEGTSMSG 540
 Db QFYDYPQCFGEVITRVEKGAAYVIGSSPNSYWGEDYVMSVGNAGVFGVQPTPEGTSMSG 540
 QY 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNTHICAHYTWAYHVLGDSVMPYRAMP 600
 Db YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNTHICAHYTWAYHVLGDSVMPYRAMP 600
 QY 601 KNTYTLPLASLPQNASYSIQASAGSVVLSKGVLYGTGVANASGAVVMTKQITENG 660
 Db KNTYTLPLASLPQNASYSIQASAGSVVLSKGVLYGTGVANASGAVVMTKQITENG 660
 QY 661 NYDVVITRNSYLPVVKIOIQGEPSPYQVSNLTATTOGQVTLKWEAPSAKKAEGSREVYK 720
 Db NYDVVITRNSYLPVVKIOIQGEPSPYQVSNLTATTOGQVTLKWEAPSAKKAEGSREVYK 720
 QY 721 RIGDGLFVITIEPANDVRANEAKVVLAAADNVWGNTGYQFLLDADHNTFGSVIPATGPLEFT 780
 Db RIGDGLFVITIEPANDVRANEAKVVLAAADNVWGNTGYQFLLDADHNTFGSVIPATGPLEFT 780
 QY 781 GTASSNLYSANFEYLIPANADPVTTQNIITVTCQGEVVPVGGVYDICIINPEPASGMWI 840
 Db GRASSNLYSANFEYLIPANADPVTTQNIITVTCQGEVVPVGGVYDICIINPEPASGMWI 840
 QY 841 AGDGGNQAPARYDDFTFEAGKKYFTTMRAGMGDGTMEVEDDPSASTYTYVYRDGTIKE 900
 Db AGDGGNQAPARYDDFTFEAGKKYFTTMRAGMGDGTMEVEDDPSASTYTYVYRDGTIKE 900
 QY 901 GLTATTFEEDGVAAGNHEVCVEKYTAGVSPKCKDVTVEGSENEFAPVQNLITGSSVGQKV 960
 Db GLTATTFEEDGVAAGNHEVCVEKYTAGVSPKCKDVTVEGSENEFAPVQNLITGSSVGQKV 960
 QY 961 TLKWDAPNGT 970
 Db TLKWDAPNGT 970

RESULT 7

AAR96028

ID AAR96028 standard; protein; 1087 AA.

AC AAR96028;

XX

DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX

DE P. gingivalis haemagglutinin hagD.

XX

KW Haemagglutinin; hagD; periodontal disease; vaccine; antibody.

XX

OS Porphyromonas gingivalis; strain FDC381.

XX

PN WO9617936-A2.

XX

PD 13-JUN-1996.

XX

PF 11-DEC-1995; 95WO-US016108.

XX

PR 09-DEC-1994; 94US-00353485.

XX

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX

XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

PI

XX WPI; 1996-287181/29.
 DR N-PSDB; AAT30652.
 XX
 PT Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 PS Claim 5; Page 65-68; 153pp; English.
 XX
 CC P. gingivalis 381 haemagglutinin hagD (AAR96028) was identified as the
 CC product of a gene (AAT30652) isolated from a P. gingivalis 318 genomic
 CC library. The haemagglutinin (see also AAR96032) can be obtd. from
 CC transformed host cells and used as a vaccine to protect humans or animals
 CC against periodontal disease. Expression in Salmonella cells allows prodn.
 CC of a live vaccine. The haemagglutinin can also be used to detect the
 CC presence of anti-P. gingivalis antibodies and to raise monoclonal
 CC antibodies for diagnostic appln. (Updated on 16-OCT-2003 to standardise
 CC OS field)
 XX
 SQ Sequence 1087 AA;

Query Match 50.9%; Score 4673.5; DB 2; Length 1087;

Best Local Similarity 81.4%; Pred. No. 3.7e-301;

Matches 900; Conservative 56; Mismatches 120; Indels 29; Gaps 10;

QY 638 GTGVANASGAVVMTKQITENGNYDVVITRNSYLPVVKIOIQGEPSPYQVSNLTATTO 697
 Db GTGVADPTVAAPVYMAKQIAENGNYDVVITRNSYLPVVKIOIQGEPSPYQVSNLTATTO 61
 QY 698 GQKVTWKWAPSAKKAEGSREVYKIDGLFVITIEPANDVRANEAKVVLAAADNVWGNTGY 757
 Db GEEVALKWDTPSAKKAEGSREVYKIDGLFVITIEPANDVRANEAKVVLAAADNVWGNTGY 121
 QY 758 QFLLDADHNTFGSVIPATGPLEFTGTASSNLYSANFEYLIPANADPVTTQNIITVTCQGEV 817
 Db QFLLDADHNTFGSVIPATGPLEFTGTASSNLYSANFEYLIPANADPVTTQNIITVTCQGEV 181
 QY 818 VIPGGVYDICIINPEPASGMWIAGDGNQAPARYDDFTFEAGKKYFTTMRAGMGDGTDM 877
 Db VIPGGVYDICIINPEPASGMWIAGDGNQAPARYDDFTFEAGKKYFTTMRAGMGDGTDM 241
 QY 878 EYEDDPSASTYTYVYRDGTIKEGLTATTFEEDGVAAGNHEVCVEKYTAGVSPKCKDV 937
 Db EYEDDPSASTYTYVYRDGTIKEGLTATTFEEDGVAAGNHEVCVEKYTAGVSPKCKDV 301
 QY 938 TVEGSENEFAPVQNLITGSSVGQKVTLKWDAPNGTTPNPNPNPNPNPNPNPNPNPNPNPN 997
 Db TVEGSENEFAPVQNLITGSSVGQKVTLKWDAPNGTTPNPNPNPNPNPNPNPNPNPNPNPN 360
 QY 998 KTIADGDGHGKPGNAPGIAGNSGCVYSFGLGGI GVLTPDNYLITPALDLDPNGGK 1057
 Db KTIADGDGHGKPGNAPGIAGNSGCVYSFGLGGI GVLTPDNYLITPALDLDPNGGK 420
 QY 1058 LTFWVCAQDANVASEHYAVVYASSTGNDASNTNALLEETITAKGVSPKAIRIGRIQGTWR 1117
 Db LTFWVCAQDANVASEHYAVVYASSTGNDASNTNALLEETITAKGVSPKAIRIGRIQGTWR 480
 QY 1118 QKTVDLPAGTKKYVAFRHFQSDTDMFYIDLDEVEIKANGKPADFTETPESSTHGEAPAEWTT 1177
 Db QKTVDLPAGTKKYVAFRHFQSDTDMFYIDLDEVEIKANGKPADFTETPESSTHGEAPAEWTT 540
 QY 1178 IDADGGQGWHLCLSSQGLDWTALHAGGSNVVSSFSWNGMALNPDNYLISKDVGTATKVXY 1237
 Db IDADGGQGWHLCLSSQGLDWTALHAGGSNVVSSFSWNGMALNPDNYLISKDVGTATKVXY 600
 QY 1238 YAVNDGFPDGHVAVMISKTGTNAGDFTVVFEEPTNGKGGARFGLSTBANGAKPOSVMI 1297
 Db YAVNDGFPDGHVAVMISKTGTNAGDFTVVFEEPTNGKGGARFGLSTBANGAKPOSVMI 660
 QY 1298 ERTVDLPAGTKKYVAFRHYNCSDINYLILDDIOFTMGGSPTPTDYTYVYRDGTIKKEGLT 1357
 Db ERTVDLPAGTKKYVAFRHYNCSDINYLILDDIOFTMGGSPTPTDYTYVYRDGTIKKEGLT 720

QY 1358 ETTTEEDGVATGHEHYCYVEVKYTAGVSKKCVYDVTVNGSTQFNPNVONLTAEQAPNSMDAIL 1417
 Db 721 ETTTEEDGVATGHEHYCYVEVKYTAGVSKKCVYDVTVNGSTQFNPNVONLTAEQAPNSMDAIL 778
 QY 1418 KWNAPAKRAEVLNEDFE-NGIPASWKTIDADGCGNNW--TTTPPPGG-----SSFAGHN 1469
 Db 779 KWEAPSGKRGELNEDFEQDAIPTGWTALDADGCGNNWDITLNEPTRGERHVLSPLRASN 838
 QY 1470 SAICVSSASHINPEFQPNPDNVLVTPPELSLPGGGTILTFWVCAQD-ANYASEHYAVYASST 1528
 Db 839 VAISYSSLLQGOEYLPFPNNFLITPKVE--GAKKITVKGSGPGLPQWSDHVALCISKS 896
 QY 1529 GNDASNFANALLEEVLTAKTAVTVAPEATRGTAQGTWYQKTVOLPAGTKYVAFRHFQCTD 1588
 Db 897 GTAAADF-----EVIFETMTYTOGGANLTREK-----DLPAGTKYVAFRHYNCTD 942
 QY 1589 FFWINLDDVIT-SGNAPSYTYTIRNNTQIASGVTTETTYRDPDLATGEYTYGVKVVYPN 1647
 Db 943 VLGIMIDDVITGEGEGPSYTYTVYRDGTKEGLETTETTYRDAGMSAQSHECYVEVKYAA 1002
 QY 1648 GESATETATLNTSLADYTAQKPYTLTVVGKTIYVTCQGEAMLYDMNGERLAAGRNTVY 1707
 Db 1003 GVSPKVCVDYIPDGADVDTAQKPYTLTVVGKTIYVTCQGEAMLYDMNGERLAAGRNTVY 1062
 QY 1708 TAQGGHYAMVVDGKSYVEKLVK 1732
 Db 1063 TAQGGHYAMVVDGKSYVEKLVK 1087
 RESULT 8
 ID AAW69486 standard; protein; 1087 AA.
 AC AAW69486;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hagD.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hagD.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN US5824791-A.
 XX
 PD 20-OCT-1998.
 XX
 PF 11-DEC-1995; 95US-00570311.
 XX
 PR 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX
 PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulskie-Fox A, Lepine G;
 XX
 DR WPI; 1998-592627/49.
 DR N-PSDB; AAV58873.
 XX
 PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 XX protease poly:peptide(s).
 XX
 PS Claim 1; Col 57-64; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hagD haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX

SQ Sequence 1087 AA;
 Query Match 50.9%; Score 4673.5; DB 2; Length 1087;
 Best Local Similarity 81.4%; Pred. No. 3.7e-301;
 Matches 900; Conservative 56; Mismatches 120; Indels 29; Gaps 10;
 QY 638 GTGVANASGVATVSMTKQITENGNYDVVITRSNLYLPIVKIQVGEPSYPQVSNLTATQ 697
 Db 2 GTTVADPTVAAPVKMAKQIAENGNYDVVMTSRNLYLPVNIQIQAGEPSYPQVNNLTAPPE 61
 QY 698 GQKVTLKWEASAKKARGSRVKRIGDGLFVTIEPANDVRANEAKVLAADNVWDNTGY 757
 Db 62 GEEVALKWDTFSAKKAEASREVKRIGDGLFVTIEPANDVRANEAKVLAADNVWDNTGY 121
 QY 758 QPLDADHNTFGSVLPATGELFTGASSNLYSANFEYLVPANADPVITTONIIVTQGBV 817
 Db 122 QPLDADHNTFGSVLPATGELFTGASSNLYSANFEYLVPANADPVITTONIIVTQGBV 181
 QY 818 VIPGVYDYCIITNPFPASGKMWIAGDGNQPARYDDFTFEAGKKYFTTMRAGMGDGTDM 877
 Db 182 VIPGVYDYCIITNPFPASGKMWIAGDGNQPARYDDFTFEAGKKYFTTMRAGMGDGTDM 241
 QY 878 EVEDDSPASYTYTVYRDGTKEGLTATTFBEDGVAAGNHEYCVBKTAGVSKVKDV 937
 Db 242 EVEDDSPASYTYTVYRDGTKEGLTATTFBEDGVAAGNHEYCVBKTAGVSKVKDV 301
 QY 938 TVEGSNEFAPVQNLGSSVGVQKVTLLKWDAPNGTNPNNPNPNPGTTLSESEFENGIPASW 997
 Db 302 TVEGSNEFAPVQNLGSSVGVQKVTLLKWDAPNGTNPNNPNPNPGTTLSESEFENGIPASW 360
 QY 998 KTIDADGSGHGWKPGNAPGAGYNSNGCVYSSFGLGIGVLTDPNYLITPALDIPNGK 1057
 Db 361 KTIDADGSGHGWKPGNAPGAGYNSNGCVYSSFGLGIGVLTDPNYLITPALDIPNGK 420
 QY 1058 LTFWVCAQDANYASEHYAVYASSTGNDASNFNALLEETITAKGVSRPAIRGTOGTWR 1117
 Db 421 LTFWVCAQDANYASEHYAVYASSTGNDASNFNALLEETITAKGVSRPAIRGTOGTWR 480
 QY 1118 OKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADETFETSESTHGEAPAEWT 1177
 Db 481 OKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADETFETSESTHGEAPAEWT 540
 QY 1178 IDADGGQGWLCISSGQDLWLTAGHGSNVVSSFWNGMALNPDNYLISKDVTGATKVXY 1237
 Db 541 IDADGGQDWLCISSGQDLWLTAGHGSNVVSSFWNGMALNPDNYLISKDVTGATKVXY 600
 QY 1238 YAVNDGFGPDHYAVMISKTGTNAGDFTVVFEETPNKGGARFGLSTEANGAKPOSVM 1297
 Db 601 YAVNDGFGPDHYAVMISKTGTNAGDFTVVFEETPNKGGARFGLSTEANGAKPOSVM 660
 QY 1298 ERTVDLPAGTKYVAFRHYNCSDLYLILDDIOFTWGGSPPTDYYTVYRDGTKEGLT 1357
 Db 661 ERTVDLPAGTKYVAFRHYNCSDLYLILDDIOFTWGGSPPTDYYTVYRDGTKEGLT 720
 QY 1358 ETTTEEDGVATGHEHYCYVEVKYTAGVSKKCVYDVTVNGSTQFNPNVONLTAEQAPNSMDAIL 1417
 Db 721 ETTTEEDGVATGHEHYCYVEVKYTAGVSKKCVYDVTVNGSTQFNPNVONLTAEQAPNSMDAIL 778
 QY 1418 KWNAPAKRAEVLNEDFE-NGIPASWKTIDADGCGNNW--TTTPPPGG-----SSFAGHN 1469
 Db 779 KWEAPSGKRGELNEDFEQDAIPTGWTALDADGCGNNWDITLNEPTRGERHVLSPLRASN 838
 QY 1470 SAICVSSASHINPEFQPNPDNVLVTPPELSLPGGGTILTFWVCAQD-ANYASEHYAVYASST 1528
 Db 839 VAISYSSLLQGOEYLPFPNNFLITPKVE--GAKKITVKGSGPGLPQWSDHVALCISKS 896
 QY 1529 GNDASNFANALLEEVLTAKTAVTVAPEATRGTAQGTWYQKTVOLPAGTKYVAFRHFQCTD 1588
 Db 897 GTAAADF-----EVIFETMTYTOGGANLTREK-----DLPAGTKYVAFRHYNCTD 942
 QY 1589 FFWINLDDVIT-SGNAPSYTYTIRNNTQIASGVTTETTYRDPDLATGEYTYGVKVVYPN 1647
 Db 943 VLGIMIDDVITGEGEGPSYTYTVYRDGTKEGLETTETTYRDAGMSAQSHECYVEVKYAA 1002
 Db 943 VLGIMIDDVITGEGEGPSYTYTVYRDGTKEGLETTETTYRDAGMSAQSHECYVEVKYAA 1002

117 ---IAPSKGIMRNEDEPKKIPYVY-KGSYQNKFFPEGEIATLDDPILRVRGVNPPAP 172
177 VQDVVANOLKVRNNEI---EVSGQADEVATQRLYDASFPYFTATKQJLFRDVTYD 233
173 LQVNPVTKLRIYETITVAVSETSEQGNILNKKGTGAG-----FEDTYKRMF-----MN 222
234 HGLDYNTPV-----RMLVVAGAKFPEALKPMLTWKAQGFYLDVHVTOBAEVGTTNASI 287
223 YBGRVTPVEEKONGRMIVIAKYEKGDIDFVDWKNQKRLRTEVKVKAEDIASPTVANAI 282
288 KAFIHKKY---NDGLAASAAPFLALVGDTOVISGE---KGKTKKVTDLIYSAVDGDF 341
283 QCFVKQXEKEGND-----LTVLLVGDHDKDIPAKITPGIKSDQV---YGVQVGNHDY 332
342 PEMYTRFMSASSPEELTNIIDKVMYKATMPDKSVLEKVLIIAGADYSWNSQVQPTIK 401
333 NEVFIGRFSCEKEDLKTQIDRTHYERNITTEDKWLQALCIAASABGGPSADNGESDIQ 392
402 Y-QMQVYVYVQEHGCTDVYNILKAPYTCY-----SHLMTGVSFANYTAHGSSETAW 450
393 HENVIANLLTQYGYTKLIK-----CYDPGVTPKNIIDAFNGGSLVNYTCHGSSETAW 444
451 ADPLLTTSOLKALNKDKYFLAIGNCITIAQFDYVQVPCFGEVITRV---KEKGAVAYIG 506
445 GTSFGCTTHVKQLTNSQNLPIFDVACVNGDFLFSMPCFAEALMRAQKQKPGTGTVAIIA 504
507 SSPNSVWGEDYVWSGANAVFGQPTFEGTSMGSYDATFLEDSYNTVNSIMWAGNLAATH 566
505 STINQSW-----ASPMRG-----QDENNEI-----LCEKH 529
567 AGNI-----GNITHIGAHYWEAYH-----VLGDGSMVPIRAMPKNTYTLPAISL 611
530 PNNIKRTFGVTVNGMFAVYKVKQGEKMLDITWTFGDPSSLVRLTLVPTKMQVTAPAQI 589
612 PQOASYSIOAS-AGSYVAISKDGLVYGTGVANASGVATVMTKQITENGNVYDVITRSN 670
590 NLTDASVNSCDYNGALATISANGKMGSAVVE-NGTATINLT-GLTNESLTLLTVVGYN 647
671 YLPIKIQIV-GEPSVQPSYNTLATTQOKVTLKWEAPSAK---KAEGSREKRLGDLG 726
648 KETVIKTINGBNPQYPSYNTLATTQOKVTLKWDAPSTKNATNTARSVDGIRELV 707
727 FVITPEAND-VRAEAKVLAADNVGDNTOGYQLDADHNTFGSVIPA-TGPIFTG-TA 783
708 LLSVSDAPELLRSQAEIVLEAHVDVNDGSGYQLLLDADHDQYQVIPSDDTLTPWNCV 767
784 SSNLYSANFEYLVPANADPVTTQNIIVTGGGEVVIPEGVYDICTNPBPASGKOWTAGD 843
768 PANLF-APFEYTVPENADPSCSPTNIMDGTASVNIIPAGTYDFAIAAPQ-ANAKIWTAGQ 825
844 GGNQPARYDDFTPEAGKKTFTMRAGMGTGDTMEVEDDPSASYTYTVYRDGTIKKEGLT 903
826 G---PTKEDDYVFEAGKKYHFLMKKSGSGDGTETLISEGGSDYTYTVYRDGTIKKEGLT 882
904 ATTFEEDGVAAGNHEYCEVVKYTAGVSPKCVKQVTVVEGSNEFAPVQNLTGSSVGQKVTLK 963
883 ATTFEEDGVATNHEYCEVVKYTAGVSPKCVKQVTVVEGSNEFAPVQNLTGSAVQKVTLK 942
964 WDAENGTPNPNPNPNPG-TTLSESFENGIPASWKIIDLADGDGHGKPGNAPGIAGNS 1022
943 WDAENGTPNPNPNPNPNPG-TTLSESFENGIPASWKIIDLADGDGHGKPGNAPGIAGNS 1002
1023 NGCVYSEFGLGGIGLVLPDNYLITPALDLPNGKLTFFWCAQDANYASHEVYASSTG 1082
1003 NGCVYSEFGLGGIGLVLPDNYLITPALDLPNGKLTFFWCAQDANYASHEVYASSTG 1062
1083 NDASNFTNALLEETITAKGVRSPKAIKRIQGTWRQKTVDLIPAGTKYVAFRHFQSTDMFY 1142
1063 NDASNFTNALLEETITAKGVRSPKAIKRIQGTWRQKTVDLIPAGTKYVAFRHFQSTDMFY 1122
1143 IDLDEVEIKANGKADFTTETTESSTHGEAPAEWTIIDADGGQGWCLLSSQLDNLTAHG 1202
1123 IDLDEVEIKANGKADFTTETTESSTHGEAPAEWTIIDADGGQGWCLLSSQLDNLTAHG 1182

1203 GSNVVSFSWNGMALNPNDYILISKDVTGATKVKYVAVNDGPPGDHYAAMI SKTGTNAGD 1262
1183 GSNVVSFSWNGMALNPNDYILISKDVTGATKVKYVAVNDGPPGDHYAAMI SKTGTNAGD 1242
1263 FTVVFEETPENGINKGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNY 1322
1243 FTVVFEETPENGINKGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNY 1302
1323 ILLDDIQTFTMGSPPTDYTYTVYRDGTIKKEGLTETTFEEDGVATNHNHEYCEVVKYTAG 1382
1303 ILLDDIQTFTMGSPPTDYTYTVYRDGTIKKEGLTETTFEEDGVATNHNHEYCEVVKYTAG 1362
1383 VSPKCKVDVTNVSSTQPNPQNLTAQAPNSMDAILKNAPASKRAE----- 1428
1363 VSPKCKVNTIPTQPNFVNKILKAQ---PDGGDVILKWEAPSAKKTGSEVKRI GDGLFV 1420
1429 -----VLNEDFENGIPASWKITIDADGNNWTTTPPPGSSPAGHNSAIC 1473
1421 TIEPANDVREANEAKVVLAAADNVGDNTOGYQL-LDADHNTFGSVIPATGFLGTASSNL 1479
1474 VSSASHINPEG--PQNPD-----NYLVT--PELSLPGSGTLTFWVCAQDANYASHEYAV 1523
1480 YSA---NFEYLIPANADPVVTTQNIIVTGGGEVVIPEG---VYDYCIITNPBPASGKMMI 1532
1524 YASSTGNDASNEFANALLBEVLTAKTVVTAPAEIRGTRAQGTWYQKTVQLPAGTKYV-APR 1582
1533 -AGDGNQPARYDDFTFE-----AGKKTFTTMR 1559
1583 HFGCTDFWINLDDVVITSGNAPSYYTYTYRNNTQIASGVTETTYRDPDLATGFTYGVK 1642
1560 RAGMGDG-----TDMVEDDPSASYTYTVYRDGTIKKEGLTETTYRDAGNSAQSHHEYCV 1614
1643 VVYPNGESALETATLNTSLADVTACKPYTLTVGKTIITVTCGGEAMIVDMGRRLLAAGR 1702
1615 VKYAAAGVSPKCVDYIIPDGVDVADTAQKPYTLTVGKTIITVTCGGEAMIVDMGRRLLAAGR 1674
1703 NTVVYTAQGHVAVMVVVDGSKSYVEKLAVK 1732
1675 NTVVYTAQGHVAVMVVVDGSKSYVEKLAVK 1704
RESULT 11
AAU08938
ID AAU08938 standard; protein; 1704 AA.
XX AAU08938;
AC AAU08938;
XX
DT 18-DEC-2001 (first entry)
XX
DE P. gingivalis high molecular weight Arg-gingipain-2.
XX
KW Periodontitis; antiinflammatory; Arg-gingipain-2; AG-2; immunogen.
XX
OS Porphyromonas gingivalis.
XX
XX
FH Key Location/Qualifiers
FT Peptide 1..227
FT Region 599..619 /label= Prepro_peptide
FT Region 670..674 /note= "Region of homology with cysteine proteases"
FT Region 719 /label= Proteolytic_component
FT Cleavage-site 720..1091
FT Protein /label= HGP_44kDa
FT /note= "Haemagglutinin protein component"
FT Cleavage-site 1091
FT Protein /label= HGP_17kDa
FT /note= "Haemagglutinin protein component"
FT Cleavage-site 1429
FT Protein /label= HGP_1704

/label= HGP 27kba
/note= "haemagglutinin protein component"

US6274718-B1.

14-AUG-2001.

25-JAN-2000; 2000US-00490931.

10-SEP-1993; 93US-00119361.

24-JUN-1994; 94US-00265441.

09-SEP-1994; 94WO-US010283.

08-NOV-1994; 94US-00336308.

(UYGE-) UNIV GEORGIA RES FOUND INC.

Travis J, Potempa JS, Barr PJ, Pavloff N;

WPI; 2001-588904/66.

N-PSDB; AAS15242.

New recombinant DNA molecule which encodes high molecular weight (mature) Arg-gingipain protein, useful for immunization against inflammation and tissue damage, comprises enzymatically active protease component and hemagglutinin component.

Claim 1; Col 29-41; 56pp; English.

The invention relates to a recombinant DNA molecule encoding high molecular weight (mature) Arg-gingipain (AG) protein, which has an enzymatically active protease component (AG-2) and a haemagglutinin component, from *P. gingivalis*. The nucleic acid is useful for producing mature Arg-gingipain protein. Immunogenic compositions comprising Arg-gingipain are useful for immunising animals including humans against inflammatory response and tissue damage caused by an archaebacterium *Porphyromonas gingivalis*, which causes progressive periodontitis. Arg-gingipain is also useful for identifying agents that modulate Arg-gingipain proteinase activity, whether by acting on the proteinase itself or preventing the interaction of the proteinase with the protein in the gingival area, such as complement factors C3 or C5. The present sequence is Arg-gingipain-2

XX Sequence 1704 AA;

Query Match 46.5%; Score 4270; DB 4; Length 1704;

Best Local Similarity 50.9%; Pred. No. 4.9e-274;

Matches 932; Conservative 204; Mismatches 456; Indels 238; Gaps 47;

6 LIIAASLLGVLYAQAQKIKLDAPTRTTCTNNSEFKQFDASFSFNEVELTKVETGGTFA 65

10 IALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTK- -VQFRMDNLKFEVQTPKGI- - 65

66 SVSIPGAFPT- - - - -GEVGSPEVPAVRKLIAPVPGATPVVRKS- -PTEQVYSLNOYG 116

66 - - - - -GOVPTYTEGVNLSEKGMFTPLISRLSLAVSDTREMKVSVSKFIEKKNVL- - - - - 116

117 SEKLMHPQMSKDDPEKVPFVYNAAYAKGFVGQELVQVIMLGMTRGVRIIAALTINP 176

117 - - - - -IASKGMWNEPDKLIPYIY- -GKSYQNKFPGGEIATLDPPFILRIVRQVNVFAP 172

177 VOYDVVANQLKVRNNIEI- - - - -EVSFQGADEVATQRLYDASFSFYFETAYKOLFNRDVIYD 233

173 LQNPVTKTLRIEITVAVSETSEQKNILNKKGTAG- - - - -FEDTYKRMF- - - - -MN 222

234 HGLDLYNTPV- - - - -RMLVAGAKFKALKPWLTKAQKGYLDVHYTDEAEVGTNNASI 287

223 YEFGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDMKNQGRILTEVKVAEDIAISPTVANAI 282

288 KAFTHKKY- - - - -NDGLAASAAPVFLALVGDVDVTSGE- - - - -KGKTKKVTDLIYSAVDGDYF 341

283 QQFVKQYEYKEGND- - - - -LITVLVGDHKOIIPAKITPGIKSDQV- - - - -YGOIVGNDHY 332

342 PEMYTERMSASSPELTNIIDKVLMEKATMPKSYLEKVLIIAGADYSWNSVQGPPTIK 401

333 NEVFIGRPSCEKEDLKQIDRTIHYERNITTEDKWLGOALCIASAECCGSPSADNGESDIQ 392

402 Y-GMYYINQEHGYTDVNNYLKAPYGCY- - - - -SHLNTGVSEFANYTAHSGSETAW 450

393 HENVIANLLTOYGYTKIK- - - - -CYDPGVTPKNIIDAFNGGISLNYTGHGSETAW 444

451 ADPLLTTSOLKALTNKDKYFLAIGNCCITAQPDYVQPCFCEVITRV- - - - -KEKGAVAYIG 506

445 GTSHEGTHVKQLTNSNQLPFDFVACVNGDFLFSPCFAEALMRAQKDGKPGTVAIIA 504

507 SSPNSYMGEDYYWSGVANAVFGVQPTFECTSMGSDATFLEDSTNTVNSIMMAGNLAATH 566

505 STINQSW- - - - -ASPMRG- - - - -QDEMNI- - - - -LCEKH 529

567 AGNI- - - - -GNITHI GAHYWEAYH- - - - -VLGDGSVMPYRAMPKNTYTLPSL 611

530 PNNIKRTFGVTMGNGFAMVEKIKDGEKMLDTWTVFGDPSLLRLVLPKMQVTAQAQI 589

612 PQNQASYSIQAS-AGSYVAISKDGLVLYGTGVANASGVATVSMTKQITENGNDVDTNRSN 670

590 NLTDASVNSCDYNGALATISANGRMFGSAVVE-NGTATINLT-GLTNSTLTTLTVVGYN 647

671 YLPVIKIQV-GEPSYQPVSNLTATTCQKQVTLKWEAPSAK- - - - -KAEGSREVRIGDGL 726

648 KETVIKTINTNGEPNPYQPVSNLTATTCQKQVTLKWDAPSTKTNTATNTARSVDGIRELV 707

727 FVTEIPAND-VRANEAKVVLADNTWGDNTGVQFLLDADHNTFGSVIPA-TGPLFTG-TA 783

708 LLSVDAPELRLSGQAEIVLEAHADVNDGSGVQIILLADHDQYGVIPSDHTFLWNCVS 767

784 SNNLSANFYLVPANADPVWTTQNIIVTQGEWVIPGVVYDICIITNPSPASGKMIAGD 843

768 PANLF-APETYVPENADESCPTNIMDGTASVNIPTAGTYDEAIAAPQ-ANAKIWIAGO 825

844 GGNQPARYDDTFEAGKYYTFMRRAGMGDGTMEVEDDSPASYTYTVVRDGTGKIKEGLT 903

826 G- - - - -PTKEDDYVEAGKYHFLMKMGSGDGTSLTISEGGSDYTYTVVRDGTGKIKEGLT 882

904 ATTFEEDGVAAGNHXYCEVVKYTAGVSPKCKDVTVEGSNEFAPVONLTGSSVGVQKVTLK 963

883 ATTFEEDGVAAGNHXYCEVVKYTAGVSPKCKDVTVEGSNEFAPVONLTGSSVGVQKVTLK 942

964 WDAPNCTPNPNPNPNPG- - - - -TTLSSEFENGIPASWKTIDADGDGHWKPGNAPGIAGYNS 1022

943 WDAPNCTPNPNPNPNPNPG- - - - -TTLSSEFENGIPASWKTIDADGDGHWKPGNAPGIAGYNS 1002

1023 NGCVYSESGLGIGVLTDPDNYLITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTG 1082

1003 NGCVYSESGLGIGVLTDPDNYLITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTG 1062

1083 NDASNFTNALLBETITAKGRVSPKAIIRGRIQGTWRKQTVLDLPAGTKYVAFRHPQSDMFY 1142

1063 NDASNFTNALLBETITAKGRVSPKAIIRGRIQGTWRKQTVLDLPAGTKYVAFRHPQSDMFY 1122

1143 IDLDEVEIKANGKRAADFTETFESSSTHGEAPAEWTTIDADGGQGMWCLSSGQDLWLTAHG 1202

1123 IDLDEVEIKANGKRAADFTETFESSSTHGEAPAEWTTIDADGGQGMWCLSSGQDLWLTAHG 1182

1203 GSNVVSFSWNGMALNPDNYLISKQVTGATKYKYIYAVNDGPGFDHYAVMI SKTGNAGD 1262

1183 GTNVASFSWNGMALNPDNYLISKQVTGATKYKYIYAVNDGPGFDHYAVMI SKTGNAGD 1242

1263 FTVVFEETNGINKGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHNVCSDLNY 1322

1243 FTVVFEETNGINKGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHNVCSDLNY 1302

1323 ILLDDIQFTMGSSPTFDYTYTVVRDGTIKI KEGLTETTFEEDGVAICNHEYCEVVKYTAG 1382

1303 ILLDDIQFTMGSSPTFDYTYTVVRDGTIKI KEGLTETTFEEDGVAICNHEYCEVVKYTAG 1362

1383 VSPKCKVDVTNSTQPNPVQNTLTAEOAPNSMDAILKWNAPASKRAE- - - - - 1428

Db 1363 VSPKCVNVTINPTQFNPVKNLKAQ--PDGDDVVLKWEAPSAKKTGSRVREKRGIDGLEFV 1420
 QY 1429 -----VLNDFENGIPASWKTIDADGDGNNWTTTPPPGGSSPAGHNSAIC 1473
 Db 1421 TIEPANDVRANEAKVLAADNVWGDNTGQFL--LDADHNTFGSVIPATGLFTGTASSML 1479
 QY 1474 VSSASHINPEG--PONPD-----NYLVT--PELSLPGGGTLPFWVCAQDANYASBHYAV 1523
 Db 1480 YSA---NPEYLLIPANADPVVTQNIIVTQGBWIPGG---VYDYCIITNPEPASCQKWI 1532
 QY 1524 YASSTGNDASNFANALLEEVULTAKTVVTAPEAIRGTRAQGTWQKTVQVLPAGTKYV-APR 1582
 Db 1533 -AGDGGNQPARYDDTFE-----AGKKYFTTMR 1559
 QY 1583 HFCTDFEWINLDDVVITSGNAPSVTYTIYRNNQTIASGVTTETTYRDPDLATGFYTYGVK 1642
 Db 1560 RAGMGDG-----TDMVEEDSPASYTYTVYRDGDKIKEGLTETTYRDAGMSAQSHCYE 1614
 QY 1643 VYYPNGESAJETATLNTISADVTAQKPYTLTVVGKTIITVTCGEAMIVDMGRRLLAAGR 1702
 Db 1615 VKYAAGVSPKVCVDYIPDGADVTAQKPYTLTVVGKTIITVTCGEAMIVDMGRRLLAAGR 1674
 QY 1703 NTVVYTAQGGHYAMVWVVDGKSVVEKLAVK 1732
 Db 1675 NTVVYTAQGGHYAMVWVVDGKSVVEKLAVK 1704

RESULT 12

ID AAR70188 standard; protein; 1704 AA.
 AC AAR70188;

XX 25-MAR-2003 (revised)
 DT 21-SEP-1995 (first entry)
 XX Arg-gingipain-2 prepolyprotein.

KW Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
 KW arginine-specific protease.

OS Porphyromonas gingivalis.

XX Key Location/Qualifiers
 FH 228..719
 FT Protein /label= Protease
 FT /note= "corresponds to Arg-gingipain-1"
 FT Region 720..1091
 FT /label= Hemagglutinin
 FT Region 1092..1429
 FT /label= Hemagglutinin
 FT Region 1430..1704
 FT /label= Hemagglutinin

XX WO9507286-A1.

XX 16-MAR-1995.

XX 09-SEP-1994; 94WO-US010283.

XX 10-SEP-1993; 93US-00119361.

PR 21-OCT-1993; 93US-00141324.

PR 24-JUN-1994; 94US-00265441.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Travis J, Potempa J, Barr PJ, Pavloff N;

XX WPI; 1995-123373/16.

DR N-PSDB; AAQ83489.

XX DNA encoding Arg-gingipain proteins - used to develop prods. for
 PT detection, treatment and prevention of periodontal disease.

XX Disclosure; Page 70-77; 89pp; English.
 XX A low mol wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50 (ATCC 53973). The sequences of the proteins were used to design PCR primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP libraries were screened with a probe based on amino acids 11-22 of the AG protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2 is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 1704 AA;
 SQ Query Match 46.5%; Score 4264; DB 2; Length 1704;
 Best Local Similarity 50.9%; Pred. No. 1.2e-273;
 Matches 931; Conservative 204; Mismatches 457; Indels 238; Gaps 47;
 QY 6 LIIAASLLGVGLYASAKIKLDAPTRTCTNNSKQFDPASFSFNEVELTKVETKGTFA 65
 Db 10 IALCSLLGGMAFAQQOTELGRNPVRLLESTQOSVTK--VQFRMDLKFTEVQTPKGI-- 65
 QY 66 SVSIPGAPPT-----GEVGSPEVPAVKLTAVPGATPVVRVKS--FTEQVYSINQVG 116
 Db 66 -----GQVETYTEGNLSEKGMPTLPILSRSLAVSDTREMKVEVSSKFIKKNVL---- 116
 QY 117 SEKLMPHPQSMKSDDEPEKVPFVYNAAYARKGVGQELTQVEMLTGMRGVRIAAALTNP 176
 Db 117 ---IAPSKGMIMRNEDPKIPYVY-GKSYSQNKFFPGEIATLDDPFIILDRVGVVNFAP 172
 QY 177 VQYDVVANOLKVRNNIEI---EVSPQGADEVATQRLYDASFPYFETAYKQIPLNRDVTYD 233
 Db 173 LQYNPVTKLRIYVEITVAVSETSEQKNILNKKGTFAG-----PETYKRMF-----NN 222
 QY 234 HGDLYNTPV-----RMLVVAGAKPEALKPWLTKAOKGVLDVHYTDEAEVGTTNASI 287
 Db 223 YEPGRYTPVEEKQNGRMIVIVAKKYEGRIDKDFWKNQKGLRTEVKKVADIASPTVANAI 282
 QY 288 KAFIHKKY---NDGLAASAAPVFLALVGDTDIVSGE--KGKTKKVTLDLYYSAVDGDYF 341
 Db 283 QQFVKQEYKEGND-----LTYVLLVGDHKDIPAKITPGIKSDQV---YGIQVGNHDY 332
 QY 342 PEMYTFRMSASSPEELTNIIDKVLMEKATMDPKSVLEKVLIIAGADYSNWSQVGTIK 401
 Db 333 NEVFIGRFSCEKEDLKTQIDRTIHYERNITTEDKWLQALCIASAEQSGPSADNGESDIQ 392
 QY 402 Y-GMQVYVYNQEHGYTDVYNVLRKAPYTGCV-----SHLNTGVSFANYTAHGSSETAW 450
 Db 393 HENVIANILTYGYTKLIK-----CYDPGVTPKNIIDAFNGGISLVNYTCHGSSETAW 444
 QY 451 ADPLLTSQLKALTNKDKYFLAIGNCCIQAQFDYVQPCFGEVITRV-----KEGAYAYIG 506
 Db 445 GTSHFGTTHVKQLTNSNQPLPFIDVACVNGDGLFSPCFEAEALMRAQKDGKPTGTVAIIA 504
 QY 507 SSPNSYMGEDYVWSVCANAVFGVQPTFEGTSMGSDATFLEDSYNTVNSIMWAGNLAATH 566
 Db 505 STINQSW-----ASPMRG-----QDMNEI-----LCEKH 529
 QY 567 AGNI-----GNITHGAHYTYEAYH-----VLGDGSMYPYRAMPKRTNYTLPASL 611
 Db 530 PNNIKETFGVTMNGMFAMVEKYKDGEXKMLDWTVFADPSLLVRLTVTPKMQVTAPAQI 589
 QY 612 PQNQASYSYQAS--AGSYVAISKDGLVLYGTGVANASGVATVSMTKQITENGNDVIVTRSN 670
 Db 590 NLTDASVNVSCDYNCAIATISANGKMFSGAVVE-NGTATINLT-GLTNESTLTTLTVVGN 647
 QY 671 YLPVIKQIQV-GEPSYPQVPSNLTATTQOKVTLKWEAPSAK---KAESREVKRIGDGL 726
 Db 648 KETVIKTINTNGEPNPYPQVPSNLTATTQOKVTLKWDAPSTKTNATNTARSVDGREL 707
 QY 727 FVTIEPAND--VRANEAKVLAADNVWGDNTGYOFLLDADHNTFGSVIPA-TGPLEFTG-TA 783
 Db 708 LLSVSDAPELLRSQAEIVLEAHDVNDGSGYQILLDADHDQYGVQVIPSDTHTLWPNCSV 767

| | | | |
|----|------|--|------|
| Qy | 234 | HGDLXNTPV-----RMLVYAGAKFEALKPMLTWKAQKGFYLDVHYHTDEAEVGTNNASI | 287 |
| Db | 223 | YEPGRYTPVEEKQNGSMIVIVAKKVBGDKDFDWMKNORGLTEVKVBAEDIASPTATANI | 282 |
| Qy | 288 | KAFITHKY---NDGLAASAAPVFLALVGDTVISGE--KGKTKKVTDLYYSAVDGVF | 341 |
| Db | 283 | QQPVKEYEKEGND-----LTYVULLVGDHKDIPAKITPGLKSDQV---YQOIVGNDHY | 332 |
| Qy | 342 | PEMYTFRMSASSPEELTNIIDKVLMYEKATMPDKSYLEKVLIIAGADYSWNVSQVGQPTIK | 401 |
| Db | 333 | NEVFIGRFSCKEDLKTQIDRTIHYERNITTEDKWLQALCITASAEGGSPSADNGESDIQ | 392 |
| Qy | 402 | Y-GMQYVYVQNEHGCTDVVNYLXAPYTCY-----SHLNTGVSFANYTAHGETAW | 450 |
| Db | 393 | HENVIANLLTQYGYTKLIK-----CUDPGVTPKNIIDAFNGGLSLVNYTGHGETAW | 444 |
| Qy | 451 | ADPLLTTSQLKALTNKDKYFLAIGNCCITAAQFDYVQVCFGEVITRV-----KEKGAYAYIG | 506 |
| Db | 445 | GTSHFGTTHVKQLTNSNQLPFIIDVACVNGDFLFSMPCFAEALMRAQKDGKGTGTVAIIA | 504 |
| Qy | 507 | SSPNSYWGEDIYVSVGNNAVFGQPTFEGTSMGSDATFLEDSYNTVNSIMMAGNLAATH | 566 |
| Db | 505 | STINQSW-----ASPMRG-----QDEMNEI-----LCEKH | 529 |
| Qy | 567 | AGNI-----GNITHIGAHYYEAYH-----VLGDGSVMYPYAMPKNTYTLPLASL | 611 |
| Db | 530 | PNNIKRTFGVWNGMFAWYKEYKDGEEKMLDWTWTFVGDPSLLVTLVTPKMQVTAQAQI | 589 |
| Qy | 612 | PQOQASYISOAS-AGSVIALSKDGLVGTGVANASGVAIVSMYTKQITENGNDVIVITRNS | 670 |
| Db | 590 | NLTDASVNSVDYNGAIIATISANGKMFSGAVVE-NGTATINLT-GLTNESTTLTLVVGYN | 647 |
| Qy | 671 | YLPVIKQIQV-GEPSYQPYNSNLTATTQGOVKVLKWEAPSAK---KAEBSREVKRIGDGL | 726 |
| Db | 648 | KETWIKIINTNGBNPPOQVNSLTATTQGOVKVLKWDAPSTKTKNATNTNARSVDGRELX | 707 |
| Qy | 727 | FVTTEPAND-VRANEAKVILAADNVWGDNTGYOFLDADHNTFGSVIPA--TGPFLTG-VA | 783 |
| Db | 708 | LLSVSDAPELLRSQAETVLEAHVWMDGSGYQIILLADHDHQYGVQVIPSOTHTLWPNCSV | 767 |
| Qy | 784 | SSNLYSANFEVLPANADPVYTONIIVTGGEVVIPGGVYDYCIITNPEPASCMMIAGD | 843 |
| Db | 768 | PANLF-APFEYTVPENADPSCPTNMLMDGTASVNI PAGTYVFAIAPAQ-ANAKIWIAGQ | 825 |
| Qy | 844 | GGNOFARYDDFTFAGKKYTFMRRAGMGDGTDMVEDDSPASYTYTVYRDGTIKKEGLT | 903 |
| Db | 826 | G---PTKEDDYVFEAGKYHFLMKMGSGDGTETLISEGGSDYTYTVYRDGTIKKEGLT | 882 |
| Qy | 904 | ATTTEEDGVAAGNHEYCVUEKYTAGVSPKVDYTVEGSNEFAPQNLTCSSVGQKVTLK | 963 |
| Db | 883 | ATTTEEDGVAATGNHEYCVUEKYTAGVSPKVDYTVEGSNEFAPQNLTCGAVGQKVTLK | 942 |
| Qy | 964 | WDAPNGTNPNNPNPNPG--FTLSSESPENGIPASWKITIDADGDGHGKMPGNAPGIAGYNS | 1022 |
| Db | 943 | WDAPNGTNPNNPNPNPGTITLSSESPENGIPASWKITIDADGDGHGKMPGNAPGIAGYNS | 1002 |
| Qy | 1023 | NGCVYSESFGGGTGVLPDPNYLTPALDLPNGKGLTFWVCAQDANYASHEHYASSTG | 1082 |
| Db | 1003 | NGCVYSESFGGGIGVLPDPNYLTPALDLPNGKGLTFWVCAQDANYASHEHYASSTG | 1062 |
| Qy | 1083 | NDASNFTWALLEETITAKGVSPKAIIRGRIQGTWRQKTVDLPAGTKVYAFRHFOSTDMFY | 1142 |
| Db | 1063 | NDASNFTWALLEETITAKGVSPKAIIRGRIQGTWRQKTVDLPAGTKVYAFRHFOSTDMFY | 1122 |
| Qy | 1143 | IDLDEVEIKANGKRADFTTEFSSTHGEBAEWTTIDADGQGWCLLSSGQLDWLTAHG | 1202 |
| Db | 1123 | IDLDEVEIKANGKRADFTTEFSSTHGEBAEWTTIDADGQGWCLLSSGQLDWLTAHG | 1182 |
| Qy | 1203 | GSNVYSSFSWNGMALNPNYLISKDVTGATKKVYYAVNDGFGDGHYAVMLSKGTGNAGD | 1262 |
| Db | 1183 | GTNVYASTFSWNGMALNPNYLISKDVTGATKKVYYAVNDGFGDGHYAVMLSKGTGNAGD | 1242 |
| Qy | 1263 | FTVWFEETPNGINKGARFGLSTEANGAKPQSVMIERTVDLPAGTKVYAFRHYNCSDLNY | 1322 |

| | | | |
|------|----|---|------|
| 1243 | Db | FTVVFETENGINKGARGFLSTEANGAKPOSWIERTVLDLPAGTKYVAFRHNCSDLNY | 1302 |
| 1323 | Qy | ILLDDTIQFTMGGSPTPTDYTYTYVRDGTIKBGLTETTFEEDGVAATGNHEYCEVVKYTAG | 1382 |
| 1303 | Db | ILLDDTIQFTMGGSPTPTDYTYTYVRDGTIKBGLTETTFEEDGVAATGNHEYCEVVKYTAG | 1362 |
| 1383 | Qy | VSPKKCVDTVTVNSQTNPVONLTABQAPNSMDAILKWNAPASKRAE----- | 1428 |
| 1363 | Db | VSPKECVNVTINPTQNPVKNLKAAQ--PDGGDVVLKWEAPSAKKTREGRVRIKGLFLV | 1420 |
| 1429 | Qy | -----VLNEDFENGIPASWKTIDADGDGNNWTTTPPPGGSSFAGHNSAIC | 1473 |
| 1421 | Db | TIFPANDVRANEAKVVLADNVWMDNTGQFL--LDADHNTFGSVIATGFLFTGTASSNL | 1479 |
| 1474 | Qy | VSSASHINFEQ--PONPD-----NYLVT--PELSLPGGGTLLTFWVCAQDANYASEHYAV | 1523 |
| 1480 | Db | YSA---NFEYLIPANADPVVTTQNIIVTQGEVWIPGG--VVDYCIITPEPASGKMWI | 1532 |
| 1524 | Qy | YASSTGNDASNFANALLEEVLTAKTUVTAPEALRGTRAQCTWQKTQVLLPAGTKYV--AFR | 1582 |
| 1533 | Db | -AGDGGNQPARYDDTFE-----AGKKYFTFMR | 1559 |
| 1583 | Qy | HFQCTDFFWLNLDDVITSGNASPYVTIYRNNTOIASGVTTETTYRDPDLATGPTYGVK | 1642 |
| 1560 | Db | RAGMGDG-----TDMVEDDSPASYTIVYRDGTKIKEGUTETTYRDAGMSAGSHEYCVE | 1614 |
| 1643 | Qy | VVYPNGESAJETATNLTSIADVTAQKPYTLTVVGKTIITVTCGEAMIYDMNGRLLAAGR | 1702 |
| 1615 | Db | VKTAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTIITVTCGEAMIYDMNGRLLAAGR | 1674 |
| 1703 | Qy | NTVYVTAQGGHYAMVVVDGKSVYEKLAVK | 1732 |
| 1675 | Db | NTVYVTAQGGHYAMVVVDGKSVYEKLAVK | 1704 |

RESULT 14
AAW69495

AAW69495
ID AAW69495 standard: protein: 1687 AA.

AA
AC
AAW69495;XX
DT 22-DEC-1998 (first entry)

XXXX

XXXX

[illegible]

05
XX
FORPBYROMONIA

PN US5824791-A.
yy

PD 20-OCT-1998.

PF 11-DEC-1995; 95US-00570311.

XX
PR 08-SEP-1988: 88US-00241640.

PR 25-JAN-1991; 91US-00647119.
PR 08-DEC-1994; 94US-00353485.

XX

PA (UABR-) UAB RES FOUND.

XX
PT
patti JM. Han N. Lantz M. Tumwasorn S. Proculske-Fox A, Lepine G;

XX
WPT: 1000 F82C27/10

DR N-PSDB; AAV58881.

Isolated *Porphyromonas gingivalis* genes - encoding haemagglutinin and/or protease polypeptide(s)).

XX
pg 167-182. English

XX

CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hagE haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease

XX Sequence 1687 AA;

Query Match 46.2%; Score 4245; DB 2; Length 1687;
 Best Local Similarity 51.6%; Pred. No. 2.2e-272;
 Matches 920; Conservative 197; Mismatches 443; Indels 224; Gaps 44;

QY 47 FSNVELNKFVEFKGTFEASVSPGAPFTEGVSPEVPAVRKLIAPVPGATPVVRKVS-- 104
 DB 30 FRMDNLKFEVTPKGMNAQVPTTEGUNISEKMPDLPILSRSLAVSDTREMKVEVVSX 89
 QY 105 FTEQVYSLNQYSGSEKMLPHQPSKSDDEPKVPFVYNAAYARKGFVGOBLTQVEMLGTM 164
 DB 90 FIEKKNVL-----IAPSGMIMRNEDPKIPVY-GKSYSQNKFPFGGEIATLDDPFIL 141
 QY 165 RGVRIAAITNPVQDVVANQLKVRNIEI---EVSFQGADEVATORLYDASFSYPETA 221
 DB 142 RDVRGQVNAFAPLQYNPVTTLRIYTEITVAVSETSEQGNILNKKGTTFAG-----FEDT 196
 QY 222 YKOLENRDVTYDHDLYNTPV-----RMLVAGAKFKALKPMLTWAKQKGFYLDVHYT 275
 DB 197 YKRMF-----MNPGRYTPVEBKQGRMIVIVAKKTEGDIKDFVWKQQRGLRTEKVA 251
 QY 276 DRAEVGTTNASIKAFIHKY---NDGLAASAPFLALVGDTDVLSGE--KGKTKKVT 329
 DB 252 EDIASPVANALIQFVKQVEYKSGND-----LTVLLVGDHDKIPAKITPGIKSDQV- 303
 QY 330 DLYSAVDGDFPEMTFTFRMSASSPELTNIIIDKVMYEKATMPDKSYLEKVLIIAGADY 399
 DB 304 --YQIIVGNHDYNEVIFGRFSCSKEDLKTQIDRTHYERNITTEDKWLQQAALCIASAE 361
 QY 390 SWNSQVQPTIKY-GMQYYNQHGVTYVNYLKAQVTCGY-----SHLTGVSVF 438
 DB 362 GPSADNGESDIQENVIANLLTOGYTKIYK-----CYDPGVTPKNIIDAFNGGISL 413
 QY 439 ANYTAHGETAWADPLLTTSQLKALNKKYFLAIGNCCITAQFDVYVQPCFGEVITRV-- 496
 DB 414 VNYTGHSETAWGTSHEFTTHVQLTNSQLPIFDVACVNGDFLFSMPCFBAALMRAQK 473
 QY 497 --KEGAYAVIGSPNSYWGEDYVWSGANAVGVQPTFEGTSMGSDATFLDEDSYNTVN 554
 DB 474 DGKPTGTVAIIASTINQSW-----ASPMRG-----QDEMNEI- 505
 QY 555 SIMWAGNLAATHAGNI---GNITHIGAHVYWEAYH-----VLGDGVSMPYRAM 599
 DB 506 -----LCEKHPNNIKRTFGVTMGMPAMVEKYKDGKEMLDWTVTFGDPSLLVRLV 558
 QY 600 PKNTVYTLPSLFPQNASYIQAS-AGSVVAISKGVLYGTGVANASGVATVSMTKOITE 658
 DB 559 PTKMQVTAQAQINLTASVNVSCDYNGAIIATISANGKFGSAVVE-NGTATINLT-GLTN 616
 QY 659 NGNYDVVITFESNYPVKIQIYV-GEPSYQPVSNLTATTOGOKVTLKWEAPSAK---KAE 714
 DB 617 ESTLITLVGYNKETVITKTINTNGEPNPQPVSNLTATTOGOKVTLKWDADPAKTNATTN 676
 QY 715 GSREVKRIGDGLFVTIRPAND-VRAEAKVVLAAADNVGDNTRYQFLLDADHNTFGSVIP 773
 DB 677 TARSVDGIRELVLLSVSDAPELLRSQAEIVLEAHDVWNDGSGVQIILLDADHQYGVIP 736
 QY 774 A-TGFLFTG-TASSNLYSANPEYLIVPANADPVVTTQNIIVTGOGEVVIPEGVYDICTNP 831
 DB 737 SDHTLWPCNSVPANLF-APPEYTVPENADPSCSPTNNIMDGTASVNIAGTYDFAIAAP 795
 QY 832 EPASGKMWIAGDGNQPARVDDFTFEAGKKYFTFMRAGMGDGTDMVEDEDDSPASYTYTV 891
 DB 796 Q-ANAKTIWIAQQG---PTKEDDYVFEAGKKYHFLMKWGGSDGTELTISBGGGSDYTYTV 851

QY 892 YRDGTKIKEGLTATTFEEDGVAAGNHEYCVVEKYTAGVSPKVKCDVTVEGSEFAFVQNL 951
 DB 852 YRDGTKIKEGLTATTFEEDGVAAGNHEYCVVEKYTAGVSPKVKCDVTVEGSEFAFVQNL 911
 QY 952 TGSSVQKVTLLKWDAPNGTNPNNPNPNPG---TTLSSSPENGIPASWKTIDADGDGHH 1008
 DB 912 TGSVQKVTLLKWDAPNGTNPNNPNPNPNPGTTLSSSPENGIPASWKTIDADGDGHH 971
 QY 1009 WKPGNAPGIAGYNSNGCVYSESFGIGGIGVLPDNYLITPALDLPNGGKLTWVCAQDAN 1068
 DB 972 WKPGNAPGIAGYNSNGCVYSESFGIGGIGVLPDNYLITPALDLPNGGKLTWVCAQDAN 1031
 QY 1069 YASEHVAVYASSTGNDASNFNTALLLETTITAKGVRSPKAIIRIGTQWTKQKTVDLPAGTK 1128
 DB 1032 YASEHVAVYASSTGNDASNFNTALLLETTITAKGVRSPKAIIRIGTQWTKQKTVDLPAGTK 1091
 QY 1129 YVAFRHQSTDMFYIDLDEVEIKANGKRAADFTTETPESSTHGEAPABWTTIDADGQGGWL 1188
 DB 1092 YVAFRHQSTDMFYIDLDEVEIKANGKRAADFTTETPESSTHGEAPABWTTIDADGQGGWL 1151
 QY 1189 CLSSGQDLWLTAHGSSNVYSSFSWNGMALNPONLYLSKDVTKATKVKYVAVYVNDGFPGDH 1248
 DB 1152 CLSSGQDLWLTAHGSSNVYSSFSWNGMALNPONLYLSKDVTKATKVKYVAVYVNDGFPGDH 1211
 QY 1249 YAMISKTCGTNAGDFTVWEETPNKNGKARFGLSTEANGAKPQSVWIERTVYDLPAGTK 1308
 DB 1212 YAMISKTCGTNAGDFTVWEETPNKNGKARFGLSTEANGAKPQSVWIERTVYDLPAGTK 1271
 QY 1309 YVAFRHNGSDLYNIIILLDDIOFTMGSSPTDYYTYVYRDGKIKEGLTETTFEEDGVAT 1368
 DB 1272 YVAFRHNGSDLYNIIILLDDIOFTMGSSPTDYYTYVYRDGKIKEGLTETTFEEDGVAT 1331
 QY 1369 GNHEYCVVEKYTAGVSPKVKCDVTVNSSTQPNVONLTAEQAPNSMDAILKWNAPASKRAE 1428
 DB 1332 GNHEYCVVEKYTAGVSPKVKCDVTVNSSTQPNVONLTAEQAPNSMDAILKWNAPASKRAE 1389
 QY 1429 -----VLNEDFENGIPASWKTIDADGDNNTTTTPP 1459
 DB 1390 GSREVKRIGDGLFVTIRPANDVRAEAKVVLAAADNVGDNTRYQFL-LOADHNTFGSVIP 1448
 QY 1460 PGSSFAGHNSAICVSSASHINPEG--PONPD-----NVLVT--PELSLPGGGITLFWV 1509
 DB 1449 ATGPLEFTGTASSNLYSA---NFEYLIPANADPVVTTQNIIVTGOGEVWIPGG---VYDY 1501
 QY 1510 CAQDANVASEHYAVYASSTGNDASNFANALBELVLTAKTVTAPEAIRGTRAQGTWYQKT 1569
 DB 1502 CITNPPASGKMWI-AGDGNQPARVDDFTFE----- 1532
 QY 1570 VQLPAGTKYV-AFRHFGCTDFFWINLDDVVITSNGNAPSYTYTYIRNNITQASGVTTTYR 1628
 DB 1533 ---AGKKYTFMRRAGMGD-----TDMVEDEDDSPASYTYTYVYRDGKIKEGLTETTYR 1583
 QY 1629 DPDLATGFTYTYGVKVVYPNGESAIEATATLNTLSADVTAGKPYTLTVVGKTIITTCOGEA 1688
 DB 1584 DAGMSAQSHHEYCVVEKYTAGVSPKVKCDYIPDGVADVTAQKPYTLTVVGKTIITTCOGEA 1643
 QY 1689 MIYDMNGRRLLAAGRNVTWYTAQGHYAVMVVVDGKSYVEKLVAK 1732
 DB 1644 MIYDMNGRRLLAAGRNVTWYTAQGHYAVMVVVDGKSYVEKLVAK 1687

RESULT 15

AAR96033

ID AAR96033 standard; protein; 1687 AA.

XX AAR96033;

DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX P. gingivalis haemagglutinin hagE.

XX Haemagglutinin; hagE; periodontal disease; vaccine; antibody.

XX OS Porphyromonas gingivalis; strain FDC381.
 XX PN MO9617936-A2.
 XX PD 13-JUN-1996.
 XX PF 11-DEC-1995; 95WO-US016108.
 XX PR 09-DEC-1994; 94US-00353485.
 XX PA (UYFL) UNIV FLORIDA.
 XX PA (UABR-) UAB RES FOUND.
 XX PI Proculus-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX WP1; 1996-287181/29.
 XX DR N-PSDB; AAT30656.
 XX PT Porphyromonas gingivalis genes and proteins - used in the detection and
 XX PT vaccination against periodontal disease.
 XX PS Claim 5; Page 138-143; 153pp; English.
 XX CC P. gingivalis 381 haemagglutinin hage (AAR96033) was identified as the
 CC product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
 CC The haemagglutinin can be obtd. from transformed host cells and used as a
 CC vaccine to protect humans or animals against periodontal disease. The
 CC Expression in Salmonella cells allows prodn. of a live vaccine. The
 CC haemagglutinin can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX SQ Sequence 1687 AA;

Query Match 46.1%; Score 4233; DB 2; Length 1687;
 Best Local Similarity 51.5%; Pred. No. 1,46-271;
 Matches 918; Conservative 197; Mismatches 445; Indels 224; Gaps 44;

QY 47 FSNFVELTKVETKGGTFASVISIGAPFTGEVSGPEVPAVKLIAPVPGATPVVRKVS-- 104
 DB 30 FRMDLKFTEVQTPKGAQVPTTEGVNLSEKGMPTLPIILSRSLAVSDTRMKVEVVSSK 89
 QY 105 FTEQVSLNQYSEKLMHPHSPMSKSDPEKVPVYNAAAARKGFPVQELTOVEMLGTM 164
 DB 90 FIEKNVL-----IAPSKGMINREDPKIPVY-GKVSQNKFPFGEIATLDDPFIL 141
 QY 165 RGVRIAALTINPVQYDVVANQLKVRNIEI---EVSFGADEVATQRLYDASFSPYPETA 221
 DB 142 RDVRGQVNVNAPLOYNPVTKRLRYTETITAVAVSETSQGKNILNKKGTFFAG-----FEDT 196
 QY 222 YKQLFNRDVTYDHGDLNTPV-----RMLVVAGAKFKEALKPLWTWKAQKGFVLDVHYT 275
 DB 197 YKRMF-----MNYEFGRYTPVEEKQNGRMIVIVAKYEGDIFKDFWNQKGLRTEVKVA 251
 QY 276 DEARVGTNTASIKAFIHKY-----NDGLAASAPVFLALVGDTDVISE--KGKTKKVT 329
 DB 252 EDIASPVTAIAIQFVQKYEKEGND-----LTYVLLVGDHDKDIPAKITPFIKSDQV- 303
 QY 330 DLYYSVGDGYFPEMYTPRMSASSPEELTNIDKVLMEKATMPDKSYLEKVLIIAGADY 389
 DB 304 --YQIVGNHDYNEVFIGRFSCSEKDLKTQIDRTIHYERNITTEDKWLGOALCIASABG 361
 QY 390 SWNSQVGPPTIKY-GMQYVYVNEQHCYTDVYNVLYKAPYTCY-----SHLNTGVSF 438
 DB 362 GPSADNGESDIOHENVIANLITQVGYTKIIL-----CYDPGVTPKNIIDAFNGGISL 413
 QY 439 ANYTAGSETAWADPLLTSTQKALTNDKDYFLAIGNOCITAQFDYVQPCFGEVITRV-- 496
 DB 414 VNYTCHGSETAWGTSHFGTTHVKQLTNSQLPFIFDVACVNGDFLFSMPCFAEALMRAQK 473
 QY 497 --KEGAVAYTGSSPNSWGEDYVWVGANAVFGVQPTFECTSMGSDATFLEDSYNTVN 554

DB 474 DGKPTGTVAIIASTINQSW-----ASPMRG-----QDENKEI- 505
 QY 555 SIWMAAGNLAATHAGNI---GNITHGAHYWEAYH-----VLGDSGVMPYRAM 599
 DB 506 -----LCEKHPNNIKRTFGVTWNGMPAMVEKYYKDGKMLDTWTVEGDSLLVRLTV 558
 QY 600 PKTNTVTLPSLPQOASYSIQAS-AGSYVAISKDGLVYGTGVANASGVATVSMTKQITE 658
 DB 559 PTMOMVTAFAQINLTADSVNVCYNGALATATISANGKMFSAVVE-NGTATINLT-GLTN 616
 QY 659 NGNYDVVITRSNVLVPIKQIQV-GEPSPYQPSNLTATTQGGKVTLKWAPSAK---KAE 714
 DB 617 ESTLTLTVVGVNKETVIKTINTNGEPNYPQPSNLTATTQGGKVTLKWAPSAKTNATTN 676
 QY 715 GSREVKRIGDGLFVTIEPAND-VRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIP 773
 DB 677 TARSVDGIRELVLLSVSDAPELLRSQAEIVLEAHDVNDGSGYQIILLDADHDQYQGVIP 736
 QY 774 A-TGPIFTG-TASSNLYSANFYLVPANADPVVTTQNIIVTGGGEVVIPIGGVYDYCITNP 831
 DB 737 SDTHTLWPNCSVPANLF-APFEYTVPENADPSCPTNMIMDGTASVNIAGTYDFAIAP 795
 QY 832 EPASGRKMTIAGDGNOPARYDDFTFEAGKYYFTMERAGMGDGTDMVEDDSDPASYTYTV 891
 DB 796 Q-ANAKIWIAGQG---PTKEDDDVFEAGKKYHFLMKKONGSGDTELTITISGGSDITYTV 851
 QY 892 YRDGTIKIKEGLTATTPEEDGVAAGNEHCVEVKYTAGVSPKVCCKDVTVEGSEFAFVQNL 951
 DB 852 YRDGTIKIKEGLTATTPEEDGVAAGNEHCVEVKYTAGVSPKVCCKDVTVEGSEFAFVQNL 911
 QY 952 TGSVQGVKVTWKWDAPNGTNPNNPNPNPG---TTLSSEFENGIPASWKTIIDADGDG 1008
 DB 912 TGSVQGVKVTWKWDAPNGTNPNNPNPNPNPG---TTLSSEFENGIPASWKTIIDADGDG 971
 QY 1009 WKDGNAPGAGYNSNGCVYSESEGLGIGVLPDNYLITPALDPLNGKLTFWVCAQDAN 1068
 DB 972 WKDGNAPGAGYNSNGCVYSESEGLGIGVLPDNYLITPALDPLNGKLTFWVCAQDAN 1031
 QY 1069 YASEHVAVYASSTGNDASFTNALLEETITAKGVRSPKAIIRIGTQRTWRTKVTDLPA 1128
 DB 1032 YASEHVAVYASSTGNDASFTNALLEETITAKGVRSPKAIIRIGTQRTWRTKVTDLPA 1091
 QY 1129 YVAFRIFQSTDMFYIIDLDEVEIKANGKRADFTFTFESSHGEAPAEWTTIIDADG 1188
 DB 1092 YVAFRIFQSTDMFYIIDLDEVEIKANGKRADFTFTFESSHGEAPAEWTTIIDADG 1151
 QY 1189 CLSSGQDLWTAGHGSNVYSSFSWNGMALPNPNYLSKDVATKVKYVAVNDGFPGDH 1248
 DB 1152 CLSSGQDLWTAGHGSNVYSSFSWNGMALPNPNYLSKDVATKVKYVAVNDGFPGDH 1211
 QY 1249 YAWISKTGNTAGDFTVVFETPENGINKGARFGLSTEANGAKPQSVIERTVDLPAGTK 1308
 DB 1212 YAWISKTGNTAGDFTVVFETPENGINKGARFGLSTEANGAKPQSVIERTVDLPAGTK 1271
 QY 1309 YVAFRHYNSDLNLYILLDDIQTMGSGPTDITYTVYRDGTIKIKEGLTFTTEEDGVAT 1368
 DB 1272 YVAFRHYNSDLNLYILLDDIQTMGSGPTDITYTVYRDGTIKIKEGLTFTTEEDGVAT 1331
 QY 1369 GNHEYCEVEVKYTAGVSPKCVDTVNSTQPNPNQNLTAEOAPNSMDAILKWNAPASKRAE 1428
 DB 1332 GNHEYCEVEVKYTAGVSPKCVDTVNSTQPNPNQNLTAEOAPNSMDAILKWNAPASKRAE 1389
 QY 1429 -----VLNEDFENGIPASWKTIIDADGNNWTTTPP 1459
 DB 1390 GSREVKRIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFL-LDADHNTFGSVIP 1448
 QY 1460 PGSSPAGHNSAICVSSASHINFEQ--PONPD-----NYLVT--PELSLPGGGTLTFWV 1509
 DB 1449 ATGPLFTGTASSNLYSA---NFEYILPANADPVVTTQNIIVTGGGEVVIPIGG---VIDY 1501
 QY 1510 CAODANYASEHYAVYASSTGNDASNPANALLEEVLTAKTAVVTAPEAIRGTRAGTWTYQKT 1569
 DB 1502 CITNPEPASKMNI-AGDGGNQPARYDDFTFE----- 1532

QY 1570 VOLPAGTKYV-AFRHFGCTDFEWINLDDWVITSGNAPSYYTYTYRNNTQIASGVTTTYR 1628
Db 1533 ----AGKKTFTMRAGMGD-----TMEVEDDSPASYTYTYVRDGTKIKEGLTETTYR 1583
QY 1629 DPDLATGFYTYGVKVVYPNGESAIEATATLNTSLADVTAKPYTLTVVGKTTITVTCQGEA 1688
Db 1584 DAGMSAQSHCYCEVKYAAAGVSPKVCVDYIPDGVADVTAGNPVTLTVVGKTTITVTCQGEA 1643
QY 1689 MIYDMNGRRLAAGRNTVVYTAQGGHYAVVWVDGKSYVEKLAVK 1732
Db 1644 MIYDMNGRRLAAGRNTVVYTAQGGHYAVVWVDGKSYVEKLAVK 1687

Search completed: May 18, 2004, 11:32:46
Job time : 93.0274 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:29:09 ; Search time 31.8582 seconds

(without alignments)
5229.534 Million cell updates/sec

Title: US-08-353-485-10

Perfect score: 9179

Sequence: 1 MRKLLLLIAASLLGVGLYLAQ.....HYAVMVVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------|--------------------|
| 1 | 9179 | 100.0 | 1732 | T30836 | lysine-specific cy |
| 2 | 4270 | 46.5 | 1704 | A54426 | gingipain R (EC 3. |
| 3 | 4217 | 45.9 | 2628 | T28651 | hemagglutinin A - |
| 4 | 3808 | 41.5 | 1526 | S49763 | gingipain R (EC 3. |
| 5 | 1222.5 | 13.3 | 991 | I40229 | arginyl endopeptid |
| 6 | 322 | 3.5 | 5291 | F90696 | hypothetical prote |
| 7 | 315 | 3.4 | 5198 | B85547 | probable RTX famil |
| 8 | 313 | 3.4 | 1993 | AF1450 | probable peptidogl |
| 9 | 300.5 | 3.3 | 2468 | A83412 | hypothetical prote |
| 10 | 295.5 | 3.2 | 4199 | S76412 | hypothetical prote |
| 11 | 280.5 | 3.1 | 3029 | S76109 | hypothetical prote |
| 12 | 276.5 | 3.0 | 2044 | AB1180 | probable peptidogl |
| 13 | 269 | 2.9 | 2167 | AF1489 | cell wall-associat |
| 14 | 265 | 2.9 | 1283 | T39174 | hypothetical serin |
| 15 | 262 | 2.9 | 2059 | D82671 | surface protein XF |
| 16 | 252 | 2.7 | 3083 | AH2493 | hypothetical prote |
| 17 | 250.5 | 2.7 | 1200 | T17404 | hyalin - sea urchi |
| 18 | 248 | 2.7 | 2554 | AB3528 | extracellular seri |
| 19 | 247.5 | 2.7 | 1939 | D97316 | probable S-layer p |
| 20 | 244 | 2.7 | 4936 | AH2515 | hypothetical prote |
| 21 | 243 | 2.6 | 3972 | S75251 | hypothetical prote |
| 22 | 242.5 | 2.6 | 2334 | S32920 | cell wall-associat |
| 23 | 241 | 2.6 | 1904 | T13256 | tail-host specific |
| 24 | 239 | 2.6 | 3624 | AD0835 | large repetitive p |
| 25 | 238 | 2.6 | 1649 | C86822 | hypothetical prote |
| 26 | 231.5 | 2.5 | 1461 | E90696 | hypothetical prote |
| 27 | 230.5 | 2.5 | 1461 | A85547 | hypothetical prote |
| 28 | 229.5 | 2.5 | 1483 | C97012 | probably celluloso |
| 29 | 229 | 2.5 | 1873 | T30944 | surface protein pr |

ALIGNMENTS

RESULT 1

T30836
lysine-specific cysteine proteinase porphyrain (EC 3.4.22.-) - Porphyromonas gingivalis
N/Alternate names: lysine-specific cysteine proteinase 1, 60K
C/Species: Porphyromonas gingivalis
C/Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text_change 17-Nov-2000
C/Accession: T30836; T30837; T30526; A53113
R/Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulski-Fox, A.; Lantz, J.
J. Bacteriol. 178, 2734-2741, 1996
A/Title: Analysis of the prtP gene encoding porphyrain, a cysteine proteinase of Porphyz
A/Reference number: Z20895; MUID:96213011; PMID:8631659
A/Accession: T30836
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1732 <BAR>
A/Cross-references: EMBL:U42210; NID:gl1314325; PID:gl1314326; PIDN:AAB06565.1
R/Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z20896
A/Accession: T30837
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A/Cross-references: EMBL:U75366; NID:92182811; PID:92182812; PIDN:AAB60809.1
R/Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A/Title: IS195, an insertion sequence-like element associated with protease genes in Por
A/Reference number: Z20844; MUID:98298016; PMID:9632563
A/Accession: T30526
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A/Cross-references: EMBL:AF017059; NID:92738802; PID:92738803; PIDN:AAC26523.1
R/Fike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A/Title: lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A/Reference number: A53113; MUID:94103245; PMID:8276827
A/Accession: A53113
A/Status: preliminary
A/Molecule type: protein
A/Residues: 229-249 <PIK>
A/Experimental source: H66
A/Note: sequence extracted from NCBI backbone (NCBIP:141690)
C/Genetics:
A/Gene: prtP; prtK
C/Keywords: cysteine proteinase; hydrolase

Query Match 100.0%; Score 9179; DB 2; Length 1732;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLLLLIAASLLGVGLYLAQSAKIKLADPTTRTCTNNNSFKQFDASFSFNEVELTKVETK 60

Db 1 MRKLLLLIISLLGVGLIYAOSAKI KLDAPTRTTCNNSEKQDFASFSNEVELTKVEYK 60
QY 61 GGTFAVSIPGAPPTGEVSGPEYPAVRKLIJAVPVGATPVVRKVSFTEQVYSLNQYSEKL 120
Db 61 GGTFAVSIPGAPPTGEVSGPEYPAVRKLIJAVPVGATPVVRKVSFTEQVYSLNQYSEKL 120
QY 121 MPHOPSMKSDDDPEKVPFVYVNAAYARKGFVGQELTQVEMLGTMRGVRIIAALTIINPVQYD 180
Db 121 MPHOPSMKSDDDPEKVPFVYVNAAYARKGFVGQELTQVEMLGTMRGVRIIAALTIINPVQYD 180
QY 181 VVANQLKVRNNIEIYVSFOGADAVATOLRYDASFSYPFETAYKOLFNRDVIYTHDGLYNT 240
Db 181 VVANQLKVRNNIEIYVSFOGADAVATOLRYDASFSYPFETAYKOLFNRDVIYTHDGLYNT 240
QY 241 FVRMLVAVAGAKFKEALKPWLITWKAQKGFYLDVHYTHDEAEVGTNNASIKAFIHKYNDGLA 300
Db 241 FVRMLVAVAGAKFKEALKPWLITWKAQKGFYLDVHYTHDEAEVGTNNASIKAFIHKYNDGLA 300
QY 301 ASAAPVFLALVGDITVLSGEGKTKKVTDLIYSAVDGDYFPEMYTFRMSASPEELTNI 360
Db 301 ASAAPVFLALVGDITVLSGEGKTKKVTDLIYSAVDGDYFPEMYTFRMSASPEELTNI 360
QY 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSWNSQVGQFTIKYGMQYIYNQEHGYTDVYNY 420
Db 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSWNSQVGQFTIKYGMQYIYNQEHGYTDVYNY 420
QY 421 LKAPYTCYSHLNTGVSFANYTAHGSTAWADPLLTTSOLKALTNKDKYFLAIGNCCITA 480
Db 421 LKAPYTCYSHLNTGVSFANYTAHGSTAWADPLLTTSOLKALTNKDKYFLAIGNCCITA 480
QY 481 QFDVQPCFGEVITRVEKGAAYAVIGSPNSYMGEDYVWSGVANAVGVQPTFEGTSMGS 540
Db 481 QFDVQPCFGEVITRVEKGAAYAVIGSPNSYMGEDYVWSGVANAVGVQPTFEGTSMGS 540
QY 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHICAHYWEAYHYVLGDSVMPYRAMP 600
Db 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHICAHYWEAYHYVLGDSVMPYRAMP 600
QY 601 KTNITVTLPSLIPONQASYSIQASAGSVYVASKDGLVGTGVANASGVATVSMTKQITFENG 660
Db 601 KTNITVTLPSLIPONQASYSIQASAGSVYVASKDGLVGTGVANASGVATVSMTKQITFENG 660
QY 661 NYDVVITRSLNLPVTKIQVGEPSYQPVSNLTATTQCKQVTLKWEAPSAKKAEGSREV 720
Db 661 NYDVVITRSLNLPVTKIQVGEPSYQPVSNLTATTQCKQVTLKWEAPSAKKAEGSREV 720
QY 721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLDADHNTFGSVIPATGELFT 780
Db 721 RIGDGLFVTIEPANDVRANEAKVLAADNVWGDNTGYQFLDADHNTFGSVIPATGELFT 780
QY 781 GTASSNLYSANFEXILV PANADPVVTTQNIIVTGGEVVIIPGVVDYCTNPEPASGKMWI 840
Db 781 GTASSNLYSANFEXILV PANADPVVTTQNIIVTGGEVVIIPGVVDYCTNPEPASGKMWI 840
QY 841 AGDGGNPARYDDTFFEAGKYYTFTMRAGMGDGTDEVEDDSDPASITYTYVYRDGTIKE 900
Db 841 AGDGGNPARYDDTFFEAGKYYTFTMRAGMGDGTDEVEDDSDPASITYTYVYRDGTIKE 900
QY 901 GLTATTTEEDGVAAGNHECYCEVYKTAGVSPKVKDVTVEGNSPEAFVQNLTGSSVGQVK 960
Db 901 GLTATTTEEDGVAAGNHECYCEVYKTAGVSPKVKDVTVEGNSPEAFVQNLTGSSVGQVK 960
QY 961 TLKWDAPNGTNPNNPNPCTTLESFENGIPASWKTIDADGGHGWKPGNAPGTAGY 1020
Db 961 TLKWDAPNGTNPNNPNPCTTLESFENGIPASWKTIDADGGHGWKPGNAPGTAGY 1020
QY 1021 NSNGCVYESFGLGIGVLTDPNLYLITPALDPLNGGKLTFWCAQDANYASEHYAVYASS 1080
Db 1021 NSNGCVYESFGLGIGVLTDPNLYLITPALDPLNGGKLTFWCAQDANYASEHYAVYASS 1080
QY 1081 TGNDASNFNALLEETITAKGVRSKPAIRGRIOGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140

Db 1081 TGNDASNFNALLEETITAKGVRSKPAIRGRIOGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
QY 1141 FYIDLDDEVIEIKANGRADFTETFESETHGEAPAEWTTTIDADGGQGLCLSSQGLDMLTA 1200
Db 1141 FYIDLDDEVIEIKANGRADFTETFESETHGEAPAEWTTTIDADGGQGLCLSSQGLDMLTA 1200
QY 1201 HGGSNVVSFSWNGMALNPNDNYLISKDVTGATKVKYKYAVNDGFFPGDHYAVMISKTGTNA 1260
Db 1201 HGGSNVVSFSWNGMALNPNDNYLISKDVTGATKVKYKYAVNDGFFPGDHYAVMISKTGTNA 1260
QY 1261 GPTVVFEETPNKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHNKCSL 1320
Db 1261 GPTVVFEETPNKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHNKCSL 1320
QY 1321 NYLLDDIIOFTMGGSPTPTDYTVYVYRDGTGKKEGLTETTFEEDGVATGNHEYCEVVKYT 1380
Db 1321 NYLLDDIIOFTMGGSPTPTDYTVYVYRDGTGKKEGLTETTFEEDGVATGNHEYCEVVKYT 1380
QY 1381 AGVSPKPKVDVTNVTQFNPNQNLTAEOAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
Db 1381 AGVSPKPKVDVTNVTQFNPNQNLTAEOAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
QY 1441 SKWTIDADGNNWTTTPPGSSFAGHNSAICVSSASHINFEQPNPDNYLTPELSIP 1500
Db 1441 SKWTIDADGNNWTTTPPGSSFAGHNSAICVSSASHINFEQPNPDNYLTPELSIP 1500
QY 1501 GGTTLTFWCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
Db 1501 GGTTLTFWCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
QY 1561 AQTWTYQKTQVLPAGTKYVAFRHFQCTDFWNLDDVITSGNAPSYYTYIYRNNTQIAS 1620
Db 1561 AQTWTYQKTQVLPAGTKYVAFRHFQCTDFWNLDDVITSGNAPSYYTYIYRNNTQIAS 1620
QY 1621 GVTETTYRDPDLATGYTYGVKVVYVNGSSAIEATLNTSLADVTAQKPEYTLTVVGKTI 1680
Db 1621 GVTETTYRDPDLATGYTYGVKVVYVNGSSAIEATLNTSLADVTAQKPEYTLTVVGKTI 1680
QY 1681 TVTCQGEAMTYDMNGRRLAAGRNTVYTAQGHYAVVWVVDGKSYVEKLAVK 1732
Db 1681 TVTCQGEAMTYDMNGRRLAAGRNTVYTAQGHYAVVWVVDGKSYVEKLAVK 1732

RESULT 2
A55426
gingipain R (BC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; I
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, J.
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain protein
A;Reference number: A55426; MUID:95138080; PMID:7836351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:9557067; PIDN:AAA69539.1; PID:g557068
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match 46.5%; Score 4270; DB 2; Length 1704;
Best Local Similarity 50.9%; Pred. No. 1.2e-228;

[illegible]

RESULT 3
T28651
hemagglutinin A - Porphyromonas gingivalis
C:Species: Porphyromonas gingivalis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C:Accession: T28651
R:Han, N.; Whitlock, J.; Proquleske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A:Title: The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381 contains
A:Reference number: Z20494; MUID:97047672; PMID:8925061
A:Accession: T28651
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2628 <HAN>
A:Cross-references: EMBL:U41807; NID: g1552410; PID: g1469916; PIDN: AAB17128.1
C:Genetics:
A:Gene: haga
Query Match 45.9%; Score 4217; DB 2; Length 2628;

Db 2318 AAGNTASSQANLLVDSGLPVDVTTINT-----VAGDDIINAA-----EAGAQDTIS 2361
QY 866 --MRRAGMGDGTOMEVEDDSPAIVTVYVYRGTGKIKEGLTATTFEEDGVAAGNHEVCVEV 923
Db 2362 GVVTRAAGDVTVVTLGGN---YITATVQ---SNLSWSVSVPADIALQALNGDLTITASV 2415
QY 924 KYTAGVSPKVCQDVTEGSENEFAPVQNLGTSSV-----GQVTLKWDAPNGTPNPENPNP 977
Db 2416 TNANGNTGSGTRDITIDANLPLGRVDTVAGDDIVNSIEHQALVI-----TGGSS 2465
QY 978 NNPPTGLSEFENGIPASWKITDADGGHGWKPG-NAPGIAGYNSNGC---VYSESFGI 1033
Db 2466 GLNAGAVLTVTI-NSWAYS-ATVQADG---SWSVGIPAAVSAWPAAGPLTVEVDGSSAN 2520
QY 1034 GGIGVLTDPNLYITPDLPLNGKLFVCAQDANYASE---HYAVYASSTGNDASNFYN 1090
Db 2521 NPVSNSHPVTVDLTAVASIN-----TVASDDVINAEKGNLTLSGSTSGIESGQTV- 2573
QY 1091 ALLEETITAKVRSKPAIRIGQGWROKTVDLPAKTKYVAFRHFQSTDMFYDLDEVEI 1150
Db 2574 -----TVTFGGKTYTASVAA--NGSW---SVNVPA-----ADLALPEGAANY 2611
QY 1151 KANGKRADETFEFESTHG---BAPAEWTIIDADGG-----QWCLCLSG 1193
Db 2612 QASVSASGNSA--SATHAYSVDASAPTLTINTIASDDIILNAEAGSPLTISGTSTAETG 2669
QY 1194 QLDWLTAGHGS---NVVSFSWNGM-----ALNPDNLYISKDVTGATKVYKYVAVNDG 1243
Db 2670 QIVTVTLNGATVGTVOAGDSVSVPTSALGALNASNTVTSATVNDK-----AGNFG 2722
QY 1244 FPGDHVAV---MISKTGNAGDFTV-----VFEETPNKNG-----GARFG 1282
Db 2723 SASHNLAVDTTAPVLTINTVAGDDIINDAEHAQALVIGTSSGEGKGVVSVVLNGKTYT 2782
QY 1283 LSTEANGAKPQSWIERTVLDLAG-----TKYVAFRHNCS- 1319
Db 2783 TTLDSAGN-----W---SVGVPAADVLTALGSAQITITASVSDRAGNSDDASRTVTVLSA 2834
QY 1320 -----LNYILLDDI-----QFTMGSGPTPTDVTYTVYRDG-----TKIREGLTETTF 1361
Db 2835 PVISINTIAGDDVINATEKGDLSLGSISDQAGTATVTLNGQVYSATTDASGWSVTV 2894
QY 1362 EDGV-ATGNHEVCVEVKYT--AGVSPKVCQDVTVNST-----QFNPVQNLTAEQAPNSMD 1414
Db 2895 PASAVSALGEATYSVTASVTNAQGSSTASHNVQVNTALPGTINPV-----ATD 2944
QY 1415 ALLKNWAPASKRAEVLNEDFENGIPASWKTIDADG-----GNWTTTTPPGSSSFA 1466
Db 2945 DII--NASEAGSAQITISGQVTAAGSTVTVBELGKTYTATVQADLSWNVSVPAADWQAL 3002
QY 1467 GHNSAICVSSASHINFEQPQNPNDLYITPELSLPGGGTILTFWVCAQDANYASEH---YAV 1523
Db 3003 G-NGELTV-NASVTNAVNGTSGTRDITIDASLPLGRVDT--VAGDDVNNIIIEHAQAVI 3058
QY 1524 YASSTGNDASNPANALLEVLTAKTIV-----TAPEAIRTRAQG 1563
Db 3059 TGSSSGFAAGTALTIVINNQTAAATVLANGWSVGVVPATDVSNWPAGTLNITVSGANGAG 3118
QY 1564 TWYQKTVOLPACTKVYAFRHFCTDFEINLDD-----VWITSNA 1604
Db 3119 TQTSITPLTVDLTAVALSMSNITSDDAINAEKGAALTLSGSTSGVBAGQTVTVTFGK 3178
QY 1605 PSYTVTIYRNNTQIASGVTTETTYRDPDLAT---GFYTYGVKVVYPNGESATETATLNTIS 1661
Db 3179 ---TVT---TTVAANGSWSTIVPAADLAALRDGDASAQVRVTVNGNSATATHEYSVDS 3231
QY 1662 LADVTAQKPYTLTVGKTIITVTCQEAEMYDNGNRRLAAGRNVTVYTAQGHYAMVVVD 1721
Db 3232 AAPT-----TINTIASDNIINASEAAAAGTVTSGTSTAQTGTLTVTLNGTNYQTTVTD 3286
QY 1722 G 1722
Db 3287 G 3287

RESULT 8

AF1450

probable peptidoglycan bound protein (LPXTG motif) lin0141 [imported] - Listeria innocua
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1450

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001

A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Me
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,
A.; Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1450

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1993 <GL>

A:Cross-references: GL:AL592022; PIDN:CAC95374.1; PID:gl6412560; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin0141

Query Match 3.4%; Score 313; DB 2; Length 1993;

Best Local Similarity 20.3%; Pred. No. 8.5e-09;

Matches 434; Conservative 208; Mismatches 702; Indels 796; Gaps 114;

QY 1 MRKLLLLIAASLLGVLYAQSAKIKLDAPTRTTCTNNSEKQFDASFSFNEVELTKVETK 60
Db 7 LRRPALLVWAAIIIG-----QLNL-----SSFRVF-----ABEK 36
QY 61 GGTFAVSIPGAFPTGEGVSGPEVPAVRKLIAPVPGATPVWRVKSTEQVSYLNQYSEKL 120
Db 37 GNEVSVEI-----QNELSKDKKAKLIKTPKNE-----EVKILTIQ----- 75
QY 121 MPHQSPMSKSDDEPKVPFYVNAAYARKGVGOELTVQVEMGLTMRGVRIAAITINPVYD 180
Db 76 -----TPDGKKVE-----GQE-----ABYT 90
QY 181 VVANQKVRNNIEVSPQGADEVATORLYDASFSYPETAYKOLFNRDVVTDHGDVNT 240
Db 91 ATKV-----GTVDFVINYQDQDK--TEKTYTASVEVNDISEVQNTPTPTVTEG-----NT 139
QY 241 PYRMLVWAGAKPKGALKPMLTWKA-----QKGFYLDVHYTDEAEVGTTNASIKAFIH 292
Db 140 PKK-----GSNLLKSAPTVMKIPDYNKTNWENGDIKDVSVT--VEFNNNNAS-----G 187
QY 293 KKYN-----DGLAASAPVFLALVG-----DTDVISGEKGGTKKVTDLVYSAVDGYFFEM 344
Db 188 KKIIFTLPDGMRFVSLPVPSSYQAPSSVDSNLS-----YFGA--GNPVGDS 232
QY 345 YTFRMSASSPELNIIDKVLMYEKAATMPDKSY-----LEKVL----- 383
Db 233 IT---SVTVPNKEG-----YNKATVGLSYELDPATEKTLNFSVQVDAAKYGYPTD 282
QY 384 -----IAGADYSNWSQVGOPTIKYGMQYVYNO--EHGYT----- 415
Db 283 LKSPIKVDAYMGEGSTPAASAEQIRAE-GKTVVGYANQKQVDTWFRNWTYSSELPEVSP 341
QY 416 -----DVVNYLKAPVT-----GCYSHLNTGVSFANYTAGSETAWA-----DPLL 455
Db 342 STATEDSYNTK-PYSVVNGISQMDGRGSKI FVPKNVKTLLTYPEGMEYGVGVVNNENSKLL 400
QY 456 TTSQKALT---NKDKYFLAIGNCCITAFQDVOP-CFG--EVITRYK-----EKGAY 502
Db 401 TNNNRITHTYPNENK-----VEIDFKQENYGIETILAVKYKVPGETEGTY 449
QY 503 AYIGSSP-----NSYWGEDIYWSVGA-----NAVFGVQPTFEGTSGMSYDATFLED 548
Db 450 ----TSPKVPHAVITTY--DDKVPETDALTNDASDTTTLAAKDTCKVVGKAANKWMLPR 503

| | | | | | | | | | |
|----|------|------------|------------------|-------------------|-----------------|----------------|---------------|----------------|------|
| QY | 549 | SY---- | NTVNSIMWAG----- | NLAATHAGNI | NIITHI | GAHYIWEAX----- | HVLGDGSV----- | 599 | |
| DB | 504 | NYIINPDRE- | SWAGLIQI | NNRRTAGYKTN | IQIYIKFEDENWEAY | TYWNPDPGT | YPGNKVKD | 562 | |
| QY | 594 | ----- | MPYB---- | AMPKTN---- | TYTLTPAS---- | LPQNCASYS | ICASAGSY----- | 627 | |
| DB | 563 | VOYKTNLSE | RYTYDGT | IFPKTN | GNKMLT | LDAVAGLQ | BGEYETE | KVANSYGVDFSVGT | 622 |
| QY | 628 | ----- | VAISKDG---- | VLYGTGVANA----- | SGVATV | SMTKQIT | TENGNDV----- | 664 | |
| DB | 623 | TAAPFKAANS | ASYGI | VKPGIT | VSQPDV | NIWNADDE | ENTKVS | GSSYTYVANNI | 682 |
| QY | 665 | VITRSNYL | PVIKQI | QVGEPS | PQYPS | NLTATT | QOQ----- | KVTLKWEA | 707 |
| DB | 683 | SPYNSG | TP-IKTARAGE----- | ----- | TVTTKASL | VMHDEY | PYGTRS | LNNPEV | 731 |
| QY | 708 | PSAKKAEGS | REKRTIGDGL | FVTIEP | ANDVRANEAK | VLAADNV | GDNTGY----- | QFLD | 763 |
| DB | 732 | GTTVKP | SSIKLTDQ | GKEVD | FVEAKT | ANGDKV | YVITKDLTV | GFEFYG | 791 |
| QY | 764 | DHNTF----- | ----- | GSVIPATGP-LF----- | ----- | TGTASS | NLYS | ANFE | 793 |
| DB | 792 | YNTTFDM | TL | SKSIHTDI | QIELLAG | SSNVI | SALGANV | FLDNL | 851 |
| QY | 794 | YLVANAD | PVVT | QNI | I | VTGQ----- | ----- | EVVIPG | 830 |
| DB | 852 | TLSPKQ | DTVA | ETFLN | VAGEIG | KAGYVEG | DDSTVS | YFTPGT | 911 |
| QY | 831 | ----- | PEPAS | KMWI | ADGG----- | NOPARYD----- | ----- | DFTFEAG | 876 |
| DB | 912 | LDIPI | PIKTOQ----- | ----- | DFGKFO | SEPPK | MDKUNGE | VPMTA | 964 |
| QY | 877 | MEVEDD | SPASY | TYTVY----- | RDGK | KI | KEGLTAT----- | TPEEDG | 917 |
| DB | 965 | DNYTSD | SIYS | NLTIDY | KANWIR | KVKTKI | SSGESQ | TIKVP | 1022 |
| QY | 918 | ----- | EYCV | EVKYTAG----- | ----- | VSPK | CKDVT | VEGSNEFA----- | 951 |
| DB | 1023 | ERDVN | PYI | AYET | NTY | SSGI | AGTRV | GAE | 1082 |
| QY | 952 | TGSSVG | QKVL- | KWDA | PNGT | PNPNPN | PGT | TLSE | 1005 |
| DB | 1083 | T----- | VELY | KWDEAT | SYET | PTKQ | QNVTT----- | TTDADG | 1120 |
| QY | 1006 | ----- | GHGW | KPGNA | PGI | AGYNS | GCVSES | FCLGG | 1061 |
| DB | 1121 | NLGIGY | KYAVK | FEKAGYKHT----- | ----- | LQNVG---- | RDK | DLNSAA----- | 1165 |
| QY | 1062 | V----- | CAQD | ANYASE | HYAV | YAS | TGNDAS | NFTN | 1116 |
| DB | 1166 | VKID | PAQ | PAQHIN | AGI | YISV | PETD----- | LKVN | 1208 |
| QY | 1117 | RQKT | VDLP | PAGTKY | VAFR | HFQ | STDMFY | IDLDE | 1171 |
| DB | 1209 | ----- | ----- | ITL | PKV----- | ----- | ----- | APTSGE | 1226 |
| QY | 1172 | -PAEW | TTID | ADG | QGW | LC | SSQLD | MLT | 1229 |
| DB | 1227 | EPDI | FQKI | QANT | DGYK | WT | TADAS | VATA | 1286 |
| QY | 1230 | GATK----- | VKYI | YAVNDG----- | ----- | FPGD | HY-A | VMIS | 1271 |
| DB | 1287 | GTEK | KSTAP | LYVT | TANG | KVAQ | KDELKI | GATN | 1346 |
| QY | 1272 | NGIN----- | ----- | KG | AREF----- | ----- | ----- | ----- | 1305 |
| DB | 1347 | NGVNS | D | AED | R | T | VV | K | 1401 |
| QY | 1306 | GTKY | VAFR | HYNC | SD | LN | VILL | DDI | 1365 |
| DB | 1402 | E----- | ----- | KDL | TEV | NH | SD | TI | 1443 |
| QY | 1366 | VATGN----- | ----- | HEYC | VB | KY | TAG | VS | 1417 |

```

Db      1444  EATGTVDTTKAGEYPVTKYN---DTTKVTITVKDDATEI-----NAHDSIT 1488
QY      1418  ----KNNAPASKRAEVLNDFENGIPASWKITDADGDNNTTTTPPPGSSFAGHNSAIC 1473
Db      1489  YTGDTWSA-----KDNFDSA-----ADRDGNE-----VA 1512
QY      1474  VSSASHINFEQPQNLYLVTPELSLPGG--TLTFWVCAQDANVASEHYAVASSTGN 1530
Db      1513  LSKVTVTNTVNTAQCTYPIT---YYGGVSKTITVTVKENKGINAHNATIYVGDSTW 1568
QY      1531  DASNFANALLEE---VLTKTQVTVTAPEAIRGTRAGQTWY-----OKTVQLPAGTKYV 1579
Db      1569  AEDNFDNAVYDKGNPVEFSKVTVTETPNVNTNKA-GTVQLKYTFDGAASKTTLTVKNIQT 1627
QY      1580  AFRHFCTDFF---W---INLDDVWITSGNAPSYYTIIYRNNTOI-ASGVTEITYRDPDL 1632
Db      1628  AVNAHNSYTVVGETWEAKNFDNSARNKDGETAFAFDEVEGNDVMTVAGTYSITYK---- 1683
QY      1633  ATGFYTYGVKVVYPNGESAIE-----TATINITSADVTAQKPYTLTVVGKTI 1680
Db      1684  YDGF-SKITKVTWKNPQTALTAHDSVVVYTGDNWSAKDNFDSAID-KAGKP-----VAYRDI 1737
QY      1681  TVTCGEAMLYDMGRRLAAGRTVTVTAQGGHYAVMVVV 1720
Db      1738  TVE---EDPTVDLN---TPGTYSVTYKYGQISKVVQITV 1770

RESULT 9
A83412
hypotheical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83412
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
:; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: A83412
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2468 <STO>
A/Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AAG05263.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA1874

Query Match
Best Local Similarity 3.3%; Score 300.5; DB 2; Length 2468;
Matches 305; Conservative 140; Mismatches 505; Indels 535; Gaps 74;

QY      534  EGTSMGSDATFLEDSYNTVNSIMWAGNLAATHAGN-----IGNITHICAH 579
Db      99  DGPLMAGY--IPQSSLAGFESLTCAGVLGMSAGTALLVGAAIGAGVAISNSSGGGG 155
QY      580  YYWEAYHVLGDSVMPYRAMPKNTYITLPASLPQNASYSIOASAGSYVAI--SKDGLVY 637
Db      156  -----GSSVPPDPTTPPKAAS---GLKIAPDGSSISGQAEAGASVGIDTNGDKPD 203
QY      638  GTGVANASGVATVSMTKOITENGNDVVIT--RSNVLPIVKIQVGEPPSPQVSNLTAT 695
Db      204  LTVIADANGNFAPLNPLTNGQTVTVVVDTPAGNASPP-AQVTAPTDTTAPATDQOVA 262
QY      696  TQGQKVTLKWEAPSAKKAEGSRVKRIGDG---LFTVIEPANDVRA-----NEAKVV 744
Db      263  PDGSSVTGKAEPGSGTVGVDTD-----GDGQPDITVVVPGGGSFEVPLNPPLTNGEITVV 316
QY      745  LAADNVWGDNTGYQFLDLADHNTFGSVIPATGFLTGTATSSNLIYSANFEYVLVPANDPVV 804
Db      317  IVTDPAGNNST--PVTVEAPDPTT--APAPATDVQVAPDGSS-----VTGNAEPGA 362

```


| | | | | | | | |
|----|------|----------|--------------------|---|------------------------------------|-------------|------|
| QY | 455 | LTTSQ--- | LKAL----- | TNKKYFLAIGNCCITAQFDVYQCFQGEVITRVKEKAYAYI | 505 | | |
| Db | 1099 | ATLYQN | SPSTKALQNVAWVNVT | SALSLK----- | TIPGQNYGNEAALITTAQQSDIYYA | 1154 | |
| QY | 506 | GSPSN | YWGEDIYWSVGANA | FGVQV----- | | 531 | |
| Db | 1155 | VLGFD | QWGHU----- | AAPIFSQPGQOKVTLVGPGCNLLAAWNLQDSDGPNVT | 1207 | | |
| QY | 532 | TEEGTS | ----- | MGSYDA----- | TFLEDSYNTVN---SIMWA----- | 559 | |
| Db | 1208 | IQLAT | NGTWTWPTILGGANAGIN | FENSELSISSINGQPAIFWTESRPPSYSLVBSQ | 1267 | | |
| QY | 560 | ----- | GNLAATHAGNIGNITHI | GAHYWEAHVHVGSGVMPYRAMPKTN | ----- | 604 | |
| Db | 1268 | NPLVLR | LGELSGTWTINQOL | SVAGNTYSTAGYTLGQ----- | VGALENTWNTGDFNP | 1322 | |
| QY | 605 | ----- | YTLPASL | PONQASYSIQASAGSYVAISKDGVLYGTGVANASGVAIVSMTKQ- | 655 | | |
| Db | 1323 | AVLFNG | GGITINSPPVPSVQ | GFSEVF---WFKLFTSDGVV---GLANLAGVFDLSUNDS | 1376 | | |
| QY | 656 | ---ITEN | GNVDV---ITRSN | YLPVLKQIQVGPSPYPQV----- | ----- | 689 | |
| Db | 1377 | LFTFL | NNGSNPQISGVITGN | WHYV----- | GTYPVKQLDLYDQLVNTLENIA | 1428 | |
| QY | 690 | ----- | SNLTAT | TQOKVTL----- | KWEAPSAK----- | KAGSREVRKIG | 723 |
| Db | 1429 | FANLP | QSGTLTLAGSGSV | LVDEFAFNYSILSVYDNGSSPSSNNNFMNLGTSQL | ----- | 1484 | |
| QY | 724 | DGLFV | TEPANDVRANEAKV | VLAAAD---NVMGDNTGYOFLDADHNTFGSVIP | ----- | A | 774 |
| Db | 1485 | NGLWG | NEVGHSHYQARFEEV | TAGPTNYSVM---DSSGNSWQSPVSNPVDVVPITLSAA | 1543 | | |
| QY | 775 | TGELFT | GTASSNLYSANFELV | PANADPVVTTQNIITWTGQGEVVI | PGVYDYCITNPEPA | 834 | |
| Db | 1544 | NNEPW | -DIVSANPAGNNAQ | IAPI-NGNPDI | FIQ-VNLTGQGSSEITG---FTVTTI | --S | 1593 |
| QY | 835 | SGKOW | IADGG----- | NOPIRYDDPTFEAGKXY----- | TFTMRRA | 869 | |
| Db | 1594 | NNQLW | TGDTGTGNAFSES | WQLGVLAAENADSTTPOLETISGDKLLNSLNP | GATFSHRVM | 1653 | |
| QY | 870 | GMGDG | DMVEDD----- | SPASVITYTVYRDG----- | ----- | 895 | |
| Db | 1654 | GATET | FTFVDTGCSPLTSPA | ---TVNIYLOGQDTPITFSLSPINQGGPVSANS | PDYLD | 1711 | |
| QY | 896 | ----- | TKIKEGLTAT | FEEDG----- | VAAGNHXY----- | 919 | |
| Db | 1712 | NQVLGI | ATIKEANDASLSLV | DGSEFVIDTNDPAIAVMASGFSNGALAYVAVGN | RGYTTQG | 1771 | |
| QY | 920 | ----- | CVEVKYTAGVSPK | CKDVTVVEGSNEFAPVQNLITGSSVGQKVTLKWDAPNGT | PNPN | 974 | |
| Db | 1772 | NAVQGS | QIILFAGG----- | DVLSQKSTLPLTTTDLSSGNGDGVLTIGTIDA | --GPDINN | 1822 | |
| QY | 975 | ----- | PNPNP | NGT---TLSESPENGI----- | ----- | PASWKT | 999 |
| Db | 1823 | VPMAL | TVGDVGDDGLVIG | NANAGTGSITYVINGHYLNGLKGKQIIDL | SNASNWS | 1882 | |
| QY | 1000 | ----- | IDADGDGHGWK | ----- | PGNAPGIAGYNSNGCVYSESFGLGIGVLTPDNYL | 1045 | |
| Db | 1883 | DQGFV | IDGVADEG-GAGFS | VAINFTGNDPQIA----- | FGAPPAKNGVAVGKVL | 1933 | |
| QY | 1046 | ITPA | ----- | LDLPN----- | GGKLTFWVCAQDA | 1067 | |
| Db | 1934 | VSPNS | QSLSPHITGHTFN | LNPQNPNQPAQTVTVGTAGYSLVSRKISGGPVTF | ----- | T | 1987 |
| QY | 1068 | NYASE | HAVYASSTCNDAS | FTNALLEETITAKGVRSKPAIRGRIQGTWRQKTVDL | PAGT | 1127 | |
| Db | 1988 | NNSGD | DLFGSSTYGVQVSN | -----QWV-GKSALP | SPNNQNYPDTTM-IAAGA | 2033 | |
| QY | 1128 | KYVAF | RHPQSDMFVIDL | DEVEIKANGKRA | DFTETFESSTHGEAP----- | ABWTT | 1177 |
| Db | 2034 | VHV | ----- | YSQTS | OPFGKVAITYTGNPIPAANGVAVNLAGAAISL | GDFT-2078 | |
| QY | 1178 | IDADG | GGMWCLSGQLD | WLTIAHGGSNVVSSFGWNGMALNPDNYLISK | QVGTAKVKVY | 1237 | |

| | | | |
|---|------|---|------|
| Db | 2079 | -----LDLDGQHQLAISA-----LGVNGSAGAV--VALSGSKFTFPSSSLQALNEAG-----N | 2124 |
| Qy | 1238 | YAVDGFPGPDHYAMISKTGTNAG----DFTVVFEETPNGINKGARFG---LSTE---- | 1286 |
| Db | 2125 | LIINGGIAGGRAGMTIMTPDGVNGDGYQDFLIAPQAGNGTQGSYLLFGFLDLSTETVPI | 2184 |
| Qy | 1287 | -----ANGAKPOSWIERTVDLPAGTKVVAFRHNCSDLNLYLLDDIQFTMGSGSPPTPD | 1340 |
| Db | 2185 | IELNAIANDSKQVFLNGSLPNOLAGTAVS-----LGNITCTQGNNEPID | 2231 |
| Qy | 1341 | -----YTYTVYR-----DGTKIKEGLTETTFEEDG-- | 1365 |
| Db | 2232 | SFLISAPNAQQFYVFGQPWLAAAGSLNLADVASDNGFVIDGNLI--GNPPTTFETTSQY | 2289 |
| Qy | 1366 | -----VATGNHEYCV-----EVKVTAGVSPKKCVDDVTNSTQ-----F | 1398 |
| Db | 2290 | IDTTPAILINGSNLYLAYKFGGNNQIFYVTNNQGSWMNSEVQLPOSAQTIPPPALAFF | 2349 |
| Qy | 1399 | NPVQNLTAEQAPNSMDAILK-----WNAP----- | 1422 |
| Db | 2350 | NNVLYLAYVDGNNGLNIITSQDQGTWNPALGGTSTPTPLFVYQGLTSLLLFAANNST | 2409 |
| Qy | 1423 | -----ASKRAEVLNED-----FENGIPASWKTIID----- | 1446 |
| Db | 2410 | STVLQFYLNSNEWIYANETGNSQTALSAISATVLGDTLYLVKGGPRNTPSTDYITST | 2469 |
| Qy | 1447 | --ADGCGNNWTTTPPGSGSFAG-----HNSAICVSSASHINFEQPGQ-- | 1486 |
| Db | 2470 | TNADLSANDSSPIPGVSSQGGPSLTNDGTNLVSLYLDSSNQLNFVSSGNGINWSSPQV | 2529 |
| Qy | 1487 | -----NPDNLYLTP-----EL-----SLPGGGLTFWVCAQDA | 1514 |
| Db | 2530 | ITNNISQPPAIAFANNELYSLVPGQSGOELNVTSPLPFTGTSILGNGSLVRL--GDV | 2587 |
| Qy | 1515 | NYASEHYA-VYASSTGNDASNFANALLEEVLTAK---TVVTAPEA-IRGTRAQGTWYQK | 1568 |
| Db | 2588 | N--GDGFADVFSGGTNAGALIFGNSTKDLTLTASGEDLVISVENATLRDVISVG----- | 2640 |
| Qy | 1569 | TVQLPAGTKVVAFRHFCTDFFWNLDDVITSGNAPSYTYTYRNTQTASGVTEYTR | 1628 |
| Db | 2641 | -----DFNGDIGIDLGLDNGNFY---VVLGNTSLG----- | 2699 |
| Qy | 1629 | DPDLATGFYTVGVKVVYPNGESAEATATLNTISLADVTAQKPYTLTVVGKTIITVTCQ-- | 1686 |
| Db | 2670 | -----DLKLSITS-----SSSPVIVNQVGGVTKSMAIGDY | 2700 |
| Qy | 1687 | -----EAMIYDMNGRRLAAGRNTVV | 1706 |
| Db | 2701 | NGDGYDDVLLMGDNGNQVAMGNSTGV | 2726 |
| RESULT 11 | | | |
| S76109 | | | |
| hypothetical protein - <i>Synechocystis</i> sp. (strain PCC 6803) | | | |
| C:Species: <i>Synechocystis</i> sp. | | | |

RESULT 11

S76109
 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
 C;Species: *Synechocystis* sp.
 A;Variety: PCC 6803
 C;date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C;Accession: S76109
 R;kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 S.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S76109
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-3029 <KAN>
 A;Cross-references: EMBL:D63989; GB:AB001339; NID:g1001396; PIDN:BAAL0087.1; PID:di01073
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 Query Match 3.1%; Score 280.5; DB 2; Length 3029;

| | |
|---|--|
| Best Local Similarity 18.9%; Pred. No. 1.1e-06; | |
| Matches 359; Conservative 219; Mismatches 705; Indels 613; Gaps 85; | |
| Qy | 20 QSAKIKLDAPTRTRTCTNNNGFKQDASFSNEVELTKVETKGGTFFASVSPGAPPTCEVG 79 |
| Db | 1135 QVTVTIDGNTYATVTTGNW-----TFNIPVADIANFEATEEVAIVS-----DLAG 1182 |
| Qy | 80 SPEVPAVRKLIADVQAT-PVVRVKSTEQYISNQVSEKLMHPHQPSMSKDDPEKVPF 138 |
| Db | 1183 NPATPATRN---ITVDITAPTVTINAI-----VDD 1210 |
| Qy | 139 VYNAARVARKGFVGOELTQVEMLCTMGV---RIAALTINPQVQDVANOLKVRNIEIE 195 |
| Db | 1211 IINAV-----EASPVAVSGTTTGVEDGQVVTIDGNTYATVGTNMTFNIPVA 1261 |
| Qy | 196 --VSFOGADEVATORLYDASPSYFETAYKQLFNRDVYTDHGLYNTP---VRMLVVG 249 |
| Db | 1262 DIANFEATEEV-----ATVS-----DLAGNEPATPATRNITV--- 1293 |
| Qy | 250 AKFKEALKPMLTWKAQGFYLDVHYTDEAEVGTNTNASIKAFIHKKYND-----GLAASA 303 |
| Db | 1294 ---DITVAPAVT-----IDTSDD-----TGAQANFI---TNDDTLVFNFGTAEAD 1332 |
| Qy | 304 APVFLALVGDST-DVISEGEEKTKKVTDLYSYAVDGDYFPEMYTFRMSASSPEELTNII 361 |
| Db | 1333 STVVSLDGIETGVTANGAGEWLDVTGTLA--DGD-----YELSVTATNPTGNSATA 1385 |
| Qy | 362 DKVLMYEKATMPDKSYLEKULLIAGADYSWNSQVQPTIKVGMQYYNQEHGYTDVNYL 421 |
| Db | 1386 TQTTIVDDTA-----PTVT-----NNAIVDDIINAV 1412 |
| Qy | 422 KAPYTCGCVSHLNTGVSPA-----NYTAHGETAWADPLLTLTSQLKALTNKOKYFL 471 |
| Db | 1413 EAGSPVAVSGTTTGVEDGQVVTIDGNTYATVGTNAWT-----FNI 1455 |
| Qy | 472 AIGNCCITAOPTYVQPCFGEVITRVKKGAVAYTGGSP-----NSY 512 |
| Db | 1456 PVAD---IANPEATE---EVVATVSD-----LAGNPATPATRNITVDTTAPTVTINAI 1502 |
| Qy | 513 WGEDYYNSVGNAVFGVQPTFEGTSMGSDATFLE-DSYN-TVNSIMWAGNLAATHAGNI 570 |
| Db | 1503 AVDDILINAVEAGSPVAVSGTTTGVEDGQVVTIDGNTYATVGTNMTFNIPVADIANF 1562 |
| Qy | 571 GNITHIGAHYWEAYHVLGDGSMVPRAMPKNTYTLPASLP----- 612 |
| Db | 1563 EATEEVA-----TVSDDLGNP--ATPATRNITVDTVAPAVTIDSIDDTCAQANDF 1612 |
| Qy | 613 ---QNOQASYSIOAGSVAISKDGLVXGTGVANASGVAIVSMTKQITENGNYDVLTRS 669 |
| Db | 1613 ITNDDTLVFNFGTAEADSTFVVVSLDGIETGVTANGAGEWILDYTGTLADGDYELSVTAT 1672 |
| Qy | 670 NYLPVFIQIQGEPSPYQPVSNLTATQGGQVTLKWEAPSAKKAEGSREKVRIGDGLFVT 729 |
| Db | 1673 N-----PTGN--SATATQIIVDDTTAPT-----VT 1695 |
| Qy | 730 IEPA---NDVRAANEAKVLAADNFWGNTGQVFLDDADHNTFGSVIAT--GPLFTGTAS 784 |
| Db | 1696 INAIIVDDIINAVEAGSEPA---VSGTTTIGE-----DGQVVTITDGNITYATVAT 1743 |
| Qy | 785 SNLVS-----ENFEYLVIPANADPVVT-----TONIIVTQGEVIVPGGVY 824 |
| Db | 1744 GNMTFNIPVADIANFE---ATEEVATVSDLAGNPATPATRNITVDTTAPTVTINAI 1799 |
| Qy | 825 DYCITNPPEAGKMWIAG-----DGGNQPARYDDFTFEA---GKKYTFTRMRAGMD-- 873 |
| Db | 1800 VDDIINAVEAGSPVAVSGTTTGVEDGQVVTIDGNTYATVGTNMTFNIPVADIANFE 1859 |
| Qy | 874 -----GTOMEVEDDPSATYTVYRVDGDKIKEGLTATTFEEDGVAAGN---HEVCV 921 |
| Db | 1860 ATEEVVATVSDLAGNPATPATRNITV---DTVAPAVTIDSIDDTGQANDFTINDDTL 1915 |
| Qy | 922 EVKTTAGVSPKCVCKDVTVEGSGNEFAPVONLTGSSVGQKV-----TLKWDAPNGTFNPEN 976 |

RESULT 12

AB1180

probable peptidoglycan bound protein (LPXTG motif) lmc0842 [imported] - *Listeria monocytogenes*
 C:Species: *Listeria monocytogenes*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AB1180
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. D.; Jones, L.M.; Karet, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madueno, E.; Maitouram, A.; Meak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; PMID:21537279; PMID:11679669

A;Accession: AB1180
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2044 <GLA>
A;Cross-references: GB:NC_003210; PIDN: CAC98920.1; PID: g16410230; GSPDB: GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0842

Query Match 3.0%; Score 276.5; DB 2; Length 2044;
Best Local Similarity 19.8%; Pred. No. 9.4e-07;
Matches 369; Conservative 211; Mismatches 688; Indels 591; Gaps 89;

QY 272 VHYTDEAEVGTNNASIKAFTHKYN-----DGLAASAPV-----FLALVGDIT 314
DB 194 VNFQDSTSTG-----KKVNFPLPGMRVSLPVSNYOAGNEDTGVLSVLGAS 242

QY 315 D-----VISGEKGTKKVLDLYSAVDGYFP---EMYTFRMSAS-----SPEELTNI 360
DB 243 DPLGIAITSEVENKETAYSQATFGTVSYELSPGTEKASFVSRVDAKYYGPTDLADP 302

QY 361 IDKVLMEKATMDPKSVLEKVLIIAGADYS-----WNSQVG---OPTIKYGMYYNQEH- 412
DB 303 IKTEVYKGEASTP-----VASADQAIRAEGNKVGVYADQNHVKTMTFTWYNDQRL 352

QY 413 -----GYTDVNYLKAPVTGCYSHLNTGVSF-----ANVTAGHSETAWADPLLTSOL 460
DB 353 SEVLASTDITDSNYTKS-----YSVVGNSLDSRGAAYVIAKNIIVTYLP-EGNEFV 406

QY 461 KALTNDKDYFLAIGNCCITAFDYVQPCFGEIVTRVKEGAYAYIGSSPNSYMGEDYYWS 520
DB 407 NVVNNAGTVLKENSNVIT--NY--PSENKVINNNKHLNSA-----TNSIYGVKYKVP 457

QY 521 VG-----ANAVFGVQPTFEGTSMGSDYATFLEDSYTV----- 553
DB 458 KGTPAGTYSTAKAPHAVITTYDGEVFSFDALTTNPNDLTTIAPLDTCVKVDITTENKMTLT 517

QY 554 -----NSIMWAGNLAAHAGNIG-----NITHIGAHYVWEAVHV----- 587
DB 518 TANGQLPNNETWAGSQIINNKSAGVKQKQYQIKFDPNWEAYWNPDPSTISGNKIS 577

QY 588 -----LPGDSVMPYRAMPKTNF--YTLPA---SLFQNOQASYSIQASAGY----- 627
DB 578 EVOYKTNLDAFRFDGALLKNNQMYRLDAKAVGLEEGEYFTEVKANVGDFAPGYQSTE 637

QY 628 -----VAISKDGVLYGT---GVANASGVAT---VSMTKQITENG-- 660
DB 638 ASATYRWNSTASYGKIIRPGVTSVQYEGAINDDADDESNTKVSIGISTYKVSABSTVANGTA 697

QY 661 -----NYDVVITRSNYPVLVPIKQIQVGEPSY-----QPVSNL 692
DB 698 AFYNKAGTKVKTSAGETINTKALVLVLDY-PYGRTRVINDPEVYLHLEGTITIQP-SSI 755

QY 693 TATTQGGK-VTLKWEAPSAKAEGRVKKR-IGDGLFV-----TIEPANDVRA 738
DB 756 KLTDQGGKDVNFTVEQETANNQDKVYALKTTDVSQVGVVGYPTKHKYLNISYDTFDTVL 815

QY 739 N-----EAKVVLADNVGDNQYQFLLDADHNTFGSVIPATGP-LFT-----CTA 783
DB 816 NKSINMDAQQVIA-----WGAN-----VTPATGNTFSDGLDWNKNGRD 855

QY 784 SSNLYSANFEYLVEPANDPVVTQNIIVTGOGEVVI-----PGGVYDYC--I 828
DB 856 NDKLLSVNSSTLSIPKQDVTVTVEFLSVAGSDKAAAMEGEDSVTSYFSPGTDADYTKVI 915

QY 829 TN-----PEPASGKMWIAGDGG-----NQPARYP-----DFTFEAGKXYTF 864
DB 916 INTSSGNASTFELVPIPKTQ-----DFGSKFQSEPFKMDKMLNGALPVSAEQDQDFV 970

QY 865 TMRAGMGDGT-NEVEDDSASY---TYTVYRDGTKIKEGLTAT-----TFEEDG 911
DB 971 SYATTATGNDYDSTDYNSYNTVADYEKVMNMVRIKVKQTQINAGETQTFKVLKVDITF--DS 1028

QY 912 VAAGNH--EYCVVEVKYTAGVSPKCKVDVTVEGSNEPAPVQNLTSSVGQKVTILKWA--- 966
DB 1029 ATEGNKIGERDINVPYRVI-----TNTFA---GSLSGTKVGAELVIGEAGML 1074

QY 967 -----PNGTPNPNPNPNPNTLSESPFENGIPASWKTITDADGDGHGKPGNAPGAGYN 1021
DB 1075 FNDKDVNGSYEKDGDEPLANETV-ELYKNVNST-----EYBPAKVGDKXNITATTD 1125

QY 1022 SNGCV---YSEFGLGGIGVLTPDNLITPALDLPNGGKLTFWVCAQDANYASEHVAVYA 1078
DB 1126 SNGKYSFSDSGVGYGNVAVKFPDK-----AGYQVTL-- 1157

QY 1079 SSTGNDASNFNALLEETITAKGVRSPKAIIRGIRIOGTWRQKTVDLPAGTKY--VAFRHQ 1136
DB 1158 KNIGKDS-----LNSDVPYSGADRGMWAKQINPTLPS-SQYTNAGYYAYN 1201

QY 1137 STMFIYDIDDEVIKANGRADFTETFESESTHGEA-----PABWTITDADGGQGWCLL 1190
DB 1202 PTQDLKVNLDKQVQV-GRSLEITLPKVASSTTQAAAEADTIEPFFKNIKATTNGYKMTVA 1260

QY 1191 SSQQLDWLTAHGSGSVSVSFWNGMALNPNYLIS-KDVTG---ATKVKYKYAVNDGFPG 1246
DB 1261 DTKAVTVQTLADGSAAVGVSTNNKTIIDVTDLTITIQDIFGTEKSKAPVYVTGTDTVA 1320

QY 1247 --DHYAVMISKGTGNAGDF-----TVVFEETPNGINKGA----- 1279
DB 1321 QQDGYTM-----GATDFSLEYKEATDLTKAQLTLAKTAAFEVKDGVNSAEDRLDQ 1373

QY 1280 -----RFGLSTEANGAKPQSVW-IERTVDLPAGTKYAFRHYNCSDLYNILLDIOFTMG 1334
DB 1374 VQVNOTQLDAIKRSGQGGVYPLTYITIKDSKTVSVTIQTVAKDLTAVNAHDSITVLGD 1433

QY 1335 SPPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCV---EVKYT--AGVSPKCV 1389
DB 1434 TWDAENF-----DSAAKREGDINAP-SDVTVTGSVNTAVAGSYPTVYTYNGSKK--I 1485

QY 1390 DVTVNSTQFNPVQNLTAECAPNS-----MDAILKWNAPASKAEVLNEDFEN 1436
DB 1486 NVTVK-----DKLTAVNAHDSITVYTDGTWAGNDFSDALDKDGNVAFADITVTGTVN 1538

QY 1437 GIPASWKTII--DADGDNWTTTPPGGSSFAGHNSAICYSSA--SHINTEGPNPNYL 1492
DB 1539 TNOAGTNTITYSDGVSKTITVTVLENKEGISAHDSITYYGVDAWNAKDNFSDAFDKDGA 1598

QY 1493 V-----TPELSLPGGCT-----LTFWVCAQDANYASEHVAVYASSTGN 1530
DB 1599 VDLVDVTVEKPIVDITTKAGAYEVYKYKVKSKITLTVAKLTAVNAHDSAIIVGTWS 1658

QY 1531 DASNFANALLEEVLTAKTVVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHFCTDFF 1590
DB 1659 AEDNFDSDLDK---GNSVAFADIEVKGTVDTD-----KVGTYPTVYTYDGVSKTI 1706

QY 1591 WINLDDVVITSGNAPSYYTYIY-----RNNQIASGVTTETTVRDPDL-----ATGFY 1637
DB 1707 NIQVKD-ILTAVNA--HDSIYIGDNWNAKDNFSAKDKDGNVWKNKDNINSENPAVDLE 1763

QY 1638 TYGV-KVYVNGBSAIEATLNTISLADVTAKPYTLT----- 1674
DB 1764 TVGVYQVITYSGG---VSKINLTVEPKTISLEVHDSMTYTDGKKAEDNFDNATDKGD 1820

QY 1675 -VWGTITVTTCQSEAMTYDMNGRLAAGRNTVVYTAQGHYAVMVVVDGSGYVEKLAIV 1731
DB 1821 QIPPKDVTVTGQ-----VDSKTAGTYEITYIYDGLKVARITV 1858

RESULT 13

AF1489
cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - I
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AF1489
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.

D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maier,
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2167 <GIA>
A:Cross-references: GH:AL592022; PIDN:CAC95686.1; PID:gl6412895; GSPDB:GNO0178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin0454

Query Match 2.9%; Score 269; DB 2; Length 2167;
Best Local Similarity 19.2%; Pred. No. 2.7e-06;
Matches 407; Conservative 251; Mismatches 741; Indels 720; Gaps 114;

| | | | |
|----|-----|--|-----|
| Qy | 69 | IPGAPPTCEGVSPEVPAVRKLIAPVVGATPVVRVK-----SFTQVYS-- | 111 |
| Db | 24 | LPGVAPLAEDSPVGVADAPDK-----KAGQAPTEVKEERTENEVFDNHDGSGFTQKIFADP | 79 |
| Qy | 112 | LN-----QYSEKLMHPQS-----MSKSDDPKVPVYNAAAYARKGF | 150 |
| Db | 80 | INMEVDGMMKIDANVEREADSDMIVPKQTPLELGFLEMED-----GAYQKUTK | 129 |
| Qy | 151 | VGQELTQVEMLGTMRGVRIALTINP-----VOYDVVANQLKVRN-----NIBIEVSF | 198 |
| Db | 130 | AGAEVT--PRLGARTGENEQAVTDQPATYKENEVYENVFRDTLRHLTFPQSVUKEDLVL | 188 |
| Qy | 199 | QGADEVAT-----QRLYDASFSYPETAYKQLENRDVYT-----DHGDL | 237 |
| Db | 189 | QEPNQVDTVVYQIETKLDLELAENGDLVLFKNKSDETMYTLPKPVMTDSNVGVTETGEAAQS | 248 |
| Qy | 238 | YNTPRMLVVAGAKPEKALK---PMLTWKQAK-----GFVLVDVHYT---DEAEVGTIN | 284 |
| Db | 249 | ENVSEFKVQLTKTVYELQKVDTEWLNDAREYPVYIDPSVRLDKIYANANINSAKPTETN | 308 |
| Qy | 285 | ASIK-----AFIHK--KYNDGLAASAAPVFLALWGTDVVIS--GEKGKTKTKV----- | 328 |
| Db | 309 | IGSKLWDSQGNAYTKLKGWDSNTGNAA--FLKM--DTSTLUNKATISKATILKVYNIWHM | 364 |
| Qy | 329 | -----TDLYSVDGDYFPPEMYTFR-----MSASSPEELTNIIDKVLN----- | 366 |
| Db | 365 | SPTVKNDLWYEAANKAKSPQVWTNTVPPVTRIGSVNVRGSEWANLDTVKTVOAWASGR | 424 |
| Qy | 367 | -----YEKATMPDKSYLEKULLIAG-----ADYSWNSQVQPIIKYGMQYIYNQEHGY | 414 |
| Db | 425 | VNNGFRLGTNINKNTYWKVVASENKNPYLEBVNTY--AQPEKPTVKTSSN--GVGTGTGY | 482 |
| Qy | 415 | TDVYNVLKAPYTCYS--HLNTGVSFANYTAHGETAWADPLLTLSQLKALTWKDKYFLAI | 473 |
| Db | 483 | MDL--SNKAVPGATSNIVISNGYKYEYFNKSTATTW-----STGKKLFFPNDE--IAK | 534 |
| Qy | 474 | GNCCITAPDYVQPCFG-----EVITRVKEKGY-----AYI-----GSSP | 509 |
| Db | 535 | G-----BFEFHDGKGTDFALDPRAQYENAFQAGSTFGLRNLTRYLFRVQAVFPQGESF | 588 |
| Qy | 510 | NSYMGEDYVWSGVANAVFGVQP-----DLVFAYMPIEKQPAVPKAYSNLAHKETGYVELNWEKSPMADGYKV | 636 |
| Db | 589 | NS----- | |
| Qy | 532 | -TFEGTSMGSDATFLEDSYNTVNSLW-----AGNLAAATHAGNIGNI-----TH | 575 |
| Db | 637 | LVFNKGAYBEYDVG--AETHWTTQNKGIWPTKEEIAEGKYALHHDGKGAELAKDPSPVYTN | 695 |
| Qy | 576 | IGAHY-----YW-----EAYHVLGDG--SVMPYRAMPKTYNTYTLPASLPQ--NQASY----- | 618 |
| Db | 696 | SGGNKYERTNYFRVYAYQAGNNTTTSVQSEPATP-----STPEALNKQLGMDVYWTSP | 750 |
| Qy | 619 | ----SIQAGSAGSVAIKSDGVLYGTGVA-----NAGSVA-----TVSMTKQITEN | 659 |

| | | |
|------|--|------|
| 751 | VRGGEVNATNGNLFHETDFNLEGRGPNVNRWTFNSQDDATGIFGKGWTSLEEKLVZE | 810 |
| 660 | GNYDVVITRSNLPVVIKIQVGEPSPYQPVSNLTATTOGQKVTLKWEAPS-----AKKA | 713 |
| 811 | ENGINVWVESD-----KKIH-----RFTTKGD-----KYEAPPGIYSEITKNA | 848 |
| 714 | EG-----SRVKRIGDGLFVTIBPANDVRANEAKVLAADNVWGDNNGYQFLLDADHN | 766 |
| 849 | DGVLKITEBKSETRFIVDG---RLKSEKOTKGE-----LTYETDGLKTS-----LRDASGR | 898 |
| 767 | TF-----GSVI-PATGP-----LFTGTASSNLSANFEYLVLPANAPVVT | 805 |
| 899 | TVTLTVEGELVKELVGEDKISYTYNDQKELISSSTARGKLYRYG-----TDGLLT | 951 |
| 806 | TQNIIVTGGEVVIPGVYVDYCIITNPEPASKWMIAGDGNQPARVDDTFEAGKYYTFT | 865 |
| 952 | S-----IYDPKHTEBKPJETTF-----AYEBEEKUTEITDPVGKKTILIS | 989 |
| 866 | MRRAGMGDGTMEVEDDSFASYTYT-----VYRDGTKIKEGLTATTFEEDGVAAGN | 916 |
| 990 | YDKA---EQQTTLTNEKKKTIYSYNDAGNPKKEIVDADGLK---LTTVYESNNLVKEV | 1044 |
| 917 | HEYCVKVTYAGVSPKVKCDVTVVEGNEFAPVON---LTGS---SVGQKVLKWDAPN--- | 968 |
| 1045 | NPKGQETAYDADGNITKATDAYGTESTYTYNDNNVTSSTDTGEKRTTYVDGADAVSE | 1104 |
| 969 | -----GTPNPNPNPNPGTTLSES-FENGIPAS-WKTIADADGGHGW | 1009 |
| 1105 | TLATESQVSSVTOYDAYGNPIRGSGELSGGNLLQNSGEKAGVSNWTLIQSDAKGSMT | 1164 |
| 1010 | --KPGNAPGIAGYNSGNCVYSESGFJGGIGVLFPDNYLITPALDLPNGKGLTFWVCAQDA | 1067 |
| 1165 | FDNTQSGAPG-----ALGGSGV-----KLTSEAMSTVK | 1192 |
| 1068 | NYAS-----EHVAVYASGTGNDASNFTN---ALL-----EETITAKGV-RSPKAIR | 1109 |
| 1193 | GYSSVTRQVDVEBETTYTTSAMTKTSGMNVADALLGLRLQDANAKOYTDAGWQSNRATS | 1252 |
| 1110 | GRIQGTW---RQKTVDLPAGTKYVAFRHFQSTDMFYID-----LDEVEIKAN | 1153 |
| 1253 | IKKNGDWVKQLTFKTSKNTQV-----LLYLNDQOPAPHKGGKGTIWDNVNQPEKG | 1303 |
| 1154 | GKRADFTETEPSS---THGEAPABWTTIDADGQCGWCLSSQDLWLTHAGSNVSVSF | 1210 |
| 1304 | SVASSNVPVNNNSFENHNGTLPTGWMR-----TGNTALTQAKVDNDSQHSQDSAV--- | 1353 |
| 1211 | SWNGMALNPONYLISKDVTGATKVKYYAVNDGFP---GDHYAVMIS-----K | 1255 |
| 1354 | -----YPERK-----ATSEAYTHIVQD-VPVYQKEAKALTISALSKSEDAKANGS | 1397 |
| 1256 | TGTVNAGDF---TVVFEETENGKKGARGLSTEANGAKPSQWIERTVDLIPAGTKYYA | 1311 |
| 1398 | VATMSNDYSVMGTYYQDGTSSVQG--QPELGT--NDMNRSAVVVKPTKPVKMKIKYVTM | 1453 |
| 1312 | FRHYNSDLNVILLDDIQFTMGGSPTPTY-----TYTVYRDGTK | 1351 |
| 1454 | FRN---GLTGKAFNFDVVRVTEGEVLTKEVYDASGNYVTASYDEBGRKISFYDIY--GNK | 1508 |
| 1352 | IKE-----GLTET-TFEEDGCVATGNHEYCVKVKYTAGVSPKKCDVTVNSTQFPNPQNILTA | 1406 |
| 1509 | TSETDEXGNKKTLYTDADNA-----LIDTKLANGTSVAYKYVDNDNGNTTE-----KNVTA | 1557 |
| 1407 | EQAPNSMDAILKWNAPASKAEVLNEDFENGIPASW---KTI---DADGPNWNWTTTPPP | 1460 |
| 1558 | S-----GKTQKNIEYVDVDNKKITAFDTALNRTIKIYEYDAAAGNETKALMPN | 1602 |
| 1461 | GGSSFAGHNSA-----ICVSSASHINPEGPQNPDNYLVTPELSPGCGTTLTFWVCAQDAN | 1515 |
| 1603 | GRVTESYDSADRMGDKWMDKLAFAEQYDPNGNQTKVTDBIN-----SIVTDKT | 1652 |
| 1516 | YASBEHAYVYASSTGNDASNFAN-----ALLBEVLTAQTVIATAPEAIRG | 1558 |
| 1653 | YDDANRITKYAERGGDSYTYKPKTKDNKGTDKVGEVAINHGXYTAKTYTYNDLDRN | 1712 |

QY 1559 TRAGTWYQKTVLPACTK--YVAFRHFQCTDFWNLDDVITSGNAPSYYTYIYRNT 1616
 Db 1713 TRVND-----GSKNAYFEDEFG-----NINVTAGNGFAANYTYDSQKVT 1754
 QY 1617 QIA-SGVTETTYRDPDLATGYTGV---KVYYPN---GESALETATLN-IT--SLADVT 1666
 Db 1755 NAAISSASGTQILDEN-----YIYDAASNETSIDNKQDGKTYEYDAVNQLTKETLPDGT 1809
 QY 1667 AQKPYLTUVVG-----KTI-----TVCQGEAMYDMNGRRJLAAGNTV 1705
 Db 1810 V-KAYTYDGRNRTQVAISGETKTIDASYNDGNQLYSWNGEALTYDANGRTSDGKYT- 1867
 QY 1706 VYTAQGGHYAVMVVDGKS 1724
 Db 1868 -YTWDTGDRUSSITKGES 1885
 RESULT 14
 T39174
 hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (Schl
 C;Species: Schizosaccharomyces pombe
 C;Date: 03-Dec-1999 #sequence revision 09-Jun-2000 #text_change 09-Jun-2000
 C;Accession: T50375; T39173; T39174; T39366
 R;Connor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1998
 A;Reference number: Z21832
 A;Accession: T50375
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1283 <CON>
 A;Cross-references: EMBL:AL021815; PIDN:CA861533.1; GSPDB:GN00067; SPDB:SPBC8B4.70C
 A;Experimental source: strain 972h; cosmid c8E4
 A;Accession: T39172
 A;Molecule type: DNA
 A;Residues: 785-1283 <CO2>
 A;Cross-references: EMBL:AL021815; PIDN:CAAL7000.1
 A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
 A;Accession: T39173
 A;Molecule type: DNA
 A;Residues: 'ME', 179, 'PLV', 183, 'W', 'KL', 556-761, 'HRGSS' <CO3>
 A;Cross-references: EMBL:AL021815; PIDN:CAAL7001.1
 A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
 A;Accession: T39174
 A;Molecule type: DNA
 A;Residues: 1-555, 'S', <CO4>
 A;Cross-references: EMBL:AL021815; PIDN:CAAL7002.1
 A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
 R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.
 submitted to the EMBL Data Library, March 1999
 A;Reference number: Z21848
 A;Accession: T39366
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-896, 'G', 898-904, 'I', 906-1283 <WOO>
 A;Cross-references: EMBL:AL035675; PIDN:CA838695.1; GSPDB:GN00067; SPDB:SPBC1289.15
 A;Experimental source: strain 972h; cosmid c1289
 C;Genetics:
 A;Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07C
 A;Map position: 2
 C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 Query Match 2.9%; Score 265; DB 2; Length 1283;
 Best Local Similarity 20.5%; Pred. No. 1.9e-06;
 Matches 298; Conservative 174; Mismatches 535; Indels 448; Gaps 71;
 QY 414 YTDVYNYLKAPYTCYS-----HLNTGVSFANTYAHGSETAWADPLLTISQLKAL 463
 Db 70 YTLTITNSQCTGNYSLTISQPLIHTNLSISKPSQTATPONT-----NITQV-SL 121
 QY 464 TNKKYFLAIGNCCITAFDYVQPCFGEVITRVKEKAYAYIGSPNSYWGEDYWSVGA 523
 Db 122 TN-----GTTNSY-----SNT 133

QY 524 NAVFGVQPTPEGTS-----MGSYDATELEDS-----YNTVNSIM- 557
 Db 134 NSL-PITDITTINGTELIIPTSYNNQSHLTIIYTSYTLNPNSTIDLSILPHSTISLST 192
 QY 558 -----WAGNLAAATHAGNIGNITHIGAHYHWEAYHVLGDGSMFYPAMPKNTNTYLPASLP 612
 Db 193 VSINDISASLSKTSPTAGTIT-----ETIVSGSV-----CVTSTFPASGT 233
 QY 613 QNQAISYIQASASYVAISKDGVLYGTGVANAGSVATVSMKQITENGNDVIVTRSNYL 672
 Db 234 TSGTVEVVEPTAGTITETIVSGSVGYTSTFPANG--TTSCTVEVVEP-----TAGTIT 284
 QY 673 PVIKQIQVGEPSYPQPVSNLTATTQGGKVTLKWEAPSAKKAEGSREVVKRIGDGLFVRIEP 732
 Db 285 ETIVSGSVGYTSTFP-----ANGTISG--TVEVVEPTAGTITETIVSGSVG---YTSTFP 334
 QY 733 ANDVRANEAKV-----LAADNYMGDNTGYQFLLDADHNTFGS--VIPATGPIFTGTASS 785
 Db 335 ANGTTSGTVEVVEPTAGTITETIVSGSVGYTSTFPASGTTSGTVEVVEPTAGTITETIVS 394
 QY 786 NLYSANFEYLVPANADPVTTQNIIVTQGEVIVPGGVYDICYTNPEPASG---KMWIAG 842
 Db 395 G--SKAFTSTFPANG--TT-----SGTVEV-----EPTAGTITKTIVSG 430
 QY 843 DGGNQPARYDDFTFEAGKKYFTMRAGMGDGTDEVEDDPSASYTYTVYRDGKIKEGL 902
 Db 431 -----SKTISTFPANGTTSGT-VEVVEPTAGTITETIVSGSV---GY 469
 QY 903 TATTFEEDGVAAGNHEYCEVKYTAGVSPKVKDVTVEGSNEFAPVQNLGSSVGQKVTL 962
 Db 470 T-STFPANGTTSGTVEV---VEPTAG---TITETIVSGSKTFTSTFPASGTTSGTVEV 521
 QY 963 KWDAPGTPNPNPNPNGTTLSESPENGIPASWKTIIDADGDGHGKPKGNAPGIAGVNS 1022
 Db 522 E-----PTAG-TITETIVSGSKAFTSTFPANGTTSGTVEVVEP-TAG--- 561
 QY 1023 NGCVYSBSFGLGIGVLT--PDNVLITPALDL--PNGGKLTFWCAQADANYASEHYAVYA 1078
 Db 562 ---TITETIVSGSVGYTSTFPASGTTSGTVEVVEPTAGTITETIVSGSVGYTSTFPASGT 618
 QY 1079 SSTGNDASNTNALLEETIT--ANGVRSKPAIRIGIQTRQKTVDPADGKTVYAFRHQ 1136
 Db 619 TSGTVEVVEPTAGTITETIVSGSVGYTSTFPASGTTSGT--VEVVEPTAGT----- 667
 QY 1137 STDYFYDLDELVEIKANGKADTFETPESSTHGEAPAEWTTIDADGQGWLCL---SSG 1193
 Db 668 -----ITETIVSGSKA-FTSTF-----PANGTT-----SGTVEVVEPTAG 701
 QY 1194 QLDWLTAHGGSNVVVSFWSNGMALNPONYLISKDVTGATKVKYVAVNDGPPGDHYAVMI 1253
 Db 702 TITETIVSGSKTFTSTFPANGTTSG-----TVEVVEPTAGTITETIVSGSVG--YTSTF 753
 QY 1254 SKGTNAGDFTVVFETFPNGINK-----GGARFGLSTEANGAKPOSVMERTVDLPAGTKY 1309
 Db 754 PASCTSGTVEV-EPTAGTITETIVSGSVFMSMTIHAHDTSSGAVIV---VEPTAGTIV- 808
 QY 1310 VAFRHYNCSDLYILLDDIOF-----TWGGS-----PTPTDYTVTVYRDGKIKEGL 1356
 Db 809 -----TETIVSGSIPFTSTIPAGTTSGTVEVVEPTAGTITETIVSGSV---GY 854
 QY 1357 TETTFEEDGVATGNHCEYCEVKYTAGVSPKVKCDVTVNSTQFNPVQNL-----AEQA 1409
 Db 855 T-STFPAGTTSGTVEV---VEPTAGTITETIVSGSVGYTSTFPAGTTSCTGTVEMVAPTA 910
 QY 1410 PMSMDAILKNAPASKAEVLNEDFENGIPAS-----FKHNSAI CVSSA-----WKTIDADGDN 1452
 Db 911 GTVTETIVSGSIP-----FTSTIPAGTTSGTVEVVEPTAGTITETIVSGSVG- 958
 QY 1453 NWTTTTPPGGSS-----FAGHNSAI CVSSA-----SHNPEGPONPDNLYTPELSL 1499
 Db 959 -YTSTFPAGTTSGTVEIVAPTAAGTITETIVSGSIPFTSTIPAGTTSG-----TVEVVE 1012


```
QY 1447 ADG-----DGNWTTTPPGSSPAGHNSAICVSSASHINPEGPQNP 1489
Db 1260 NDGVKGVSDVALGTTGLTIANGPAVTASGDAGSKVISHVAGAVSETSDAVNGSQ--- 1316
QY 1490 NYLVTPELSLPGGGTLTFWVCAQDANYASHYAVYASSTGNDASNPFANALLEEVLTAKT 1549
Db 1317 LNAVQVOASQP---VTF-----TGNEGA-----VKESLGQSVV 1346
QY 1550 VTAPEAIRGTRAGQWVQKTVQLPAGTKYVAFRHFCTDFFWINLDDVVTSGNAPSITY 1609
Db 1347 ISGESSTAGTYSGGN-LKSVVDEAAGT-----IHLQADSPKFGN 1385
QY 1610 TIYRNNTQIASGVTTETTYRDPDLATGFTYTVGVKVVYPNGESAETATLNTLSLADVT--A 1667
Db 1386 VVINNGGKI-SGVTAGTE-----ETDAVNFSQLKSISTAV 1419
QY 1668 QKPYTLTVVGK-----TITVTCQGEA-----MIYDMNGRRLAAGR- 1702
Db 1420 DQGWTLTASGANGSKVASGGTVDLKNTDGNLTISKSGDSNDVVFNLSLDLKEKSIIVGNT 1479
QY 1703 -----NTVYTAQGHYAMVWVDGKSYVEKLV 1731
Db 1480 QLDKDGKVVSSNVLLDSNELVITSHSSTSSVKTLANGESVYVNRVV 1525
```

Search completed: May 18, 2004, 11:36:13
Job time : 48.8582 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:22:23 ; Search time 21.7568 Seconds
(without alignments)
4145:156 Million cell updates/sec

Title: US-08-353-485-10
Perfect score: 9179
Sequence: 1 MRKLLLLIAALLGLVGLAQ.....HYAVMVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|--------------|--------------------|
| 1 | 4217 | 45.9 | 2628 | 1 HGA2_PORGI | Q51845 porphyromon |
| 2 | 4200.5 | 45.8 | 2164 | 1 HGAI_PORGI | P59915 porphyromon |
| 3 | 1946.5 | 21.2 | 989 | 1 PRTH_PORGI | P46071 porphyromon |
| 4 | 1222.5 | 13.3 | 991 | 1 CPGL_PORGI | P28784 porphyromon |
| 5 | 541 | 5.9 | 736 | 1 CPGL_PORGI | P95493 porphyromon |
| 6 | 250.5 | 2.7 | 1200 | 1 HVAL_STRPU | Q76536 strongyloce |
| 7 | 242.5 | 2.6 | 2334 | 1 WAPA_BACSU | Q07833 bacillus su |
| 8 | 224 | 2.4 | 2894 | 1 YD96_METJA | Q58791 methanococc |
| 9 | 223 | 2.4 | 1541 | 1 IGAI_HABIN | P42782 haemophilus |
| 10 | 220.5 | 2.4 | 1902 | 1 P2P_LACIC | P15293 lactococcus |
| 11 | 219.5 | 2.4 | 1953 | 1 BIGA_SALTY | P25927 salmonella |
| 12 | 218.5 | 2.4 | 3178 | 1 YS89_CAEEL | Q09624 caenorhabdi |
| 13 | 214 | 2.3 | 1861 | 1 APU_THERTU | P38536 t amylolupl |
| 14 | 211 | 2.3 | 2358 | 1 YEEJ_ECOLI | P76347 escherichia |
| 15 | 209.5 | 2.3 | 1655 | 1 OMPB_RICCN | Q9kka3 r outer mem |
| 16 | 209 | 2.3 | 1090 | 1 GUXB_CELFI | P50899 cellulomona |
| 17 | 209 | 2.3 | 2660 | 1 YEEJ_ECO57 | Q8x8v7 escherichia |
| 18 | 208.5 | 2.3 | 1902 | 1 P1P_LACIC | P16271 lactococcus |
| 19 | 202.5 | 2.2 | 1481 | 1 APU_THERT | P38939 t amylolupl |
| 20 | 202.5 | 2.2 | 1656 | 1 OMPB_RICJA | Q06653 r outer mem |
| 21 | 202.5 | 2.2 | 1848 | 1 CBPA_CLOCL | P38058 clostridium |
| 22 | 200.5 | 2.2 | 1654 | 1 OMPB_RICRI | Q53047 r outer mem |
| 23 | 199 | 2.2 | 1645 | 1 OMPB_RICTY | P96989 r outer mem |
| 24 | 196.5 | 2.1 | 1902 | 1 P1P_LACIC | P15282 lactococcus |
| 25 | 194 | 2.1 | 1279 | 1 APU_THESA | P36905 t amylolupl |
| 26 | 194 | 2.1 | 2249 | 1 OMPA_RICRI | P15921 rickettsia |
| 27 | 192 | 2.1 | 1475 | 1 APU_THERT | P16950 t amylolupl |
| 28 | 187 | 2.0 | 1243 | 1 VG37_BP3 | Q38394 bacterioph |
| 29 | 187 | 2.0 | 1256 | 1 ATL_STAAR | P52081 staphylococ |
| 30 | 185 | 2.0 | 1902 | 1 P2P_LACPA | Q02470 lactobacill |
| 31 | 184.5 | 2.0 | 1770 | 1 EMPC_CHLTR | Q84419 chlamydia t |
| 32 | 184 | 2.0 | 1450 | 1 MP5F_CHICK | Q02173 gallus gall |
| 33 | 183.5 | 2.0 | 1260 | 1 ALSI_CANAL | P46590 candida alb |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 183.5 | 2.0 | 1577 | 1 HLYA_PROMI | P16466 proteus mir |
| 35 | 183 | 2.0 | 2021 | 1 OMPA_RICCN | Q52657 rickettsia |
| 36 | 181 | 2.0 | 4590 | 1 FATH_HUMAN | Q14517 homo sapien |
| 37 | 179.5 | 2.0 | 1300 | 1 L20K_RICRI | P14914 rickettsia |
| 38 | 177.5 | 1.9 | 1325 | 1 YDEK_ECOLI | P32051 escherichia |
| 39 | 175.5 | 1.9 | 1341 | 1 VG37_BPT2 | P07067 bacterioph |
| 40 | 175.5 | 1.9 | 1829 | 1 FRPC_NEIMB | Q9jyv5 neisseria m |
| 41 | 175.5 | 1.9 | 4303 | 1 PKD1_HUMAN | P98161 homo sapien |
| 42 | 175 | 1.9 | 6632 | 1 UN89_CAEEL | O01761 caenorhabdi |
| 43 | 174.5 | 1.9 | 872 | 1 GUXA_CELFI | P50401 cellulomona |
| 44 | 174.5 | 1.9 | 1419 | 1 ALAI_CANAL | Q13368 candida alb |
| 45 | 174 | 1.9 | 1045 | 1 GUNB_CELFI | P26225 cellulomona |

ALIGNMENTS

RESULT 1
HGA2_PORGI
ID HGA2_PORGI STANDARD; PRT; 2628 AA.
AC Q51845;
DT 20-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin A precursor.
GN HGA
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97047672; PubMed=8926061;
RA Han N., Whitlock J., Proguelske-Fox A.;
RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
contains four large, contiguous, direct repeats.";
RL Infect. Immun. 64:4000-4007(1996).
CC -!- FUNCTION: Agglutinates erythrocytes.
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U41807; AAB17128.1; -;
DR PIR; T28651; T28651.
KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 45.9%; Score 4217; DB 1; Length 2628;
Best Local Similarity 51.0%; Pred. No. 4, 1e-223;
Matches 958; Conservative 114; Mismatches 363; Indels 442; Gaps 53;

| | | | |
|----|-----|--|-----|
| QY | 179 | YDVANQLKVRNNIETIEVSFQ-----GADEVATORLYDASFPYFETAYKQLFNDRVYT | 232 |
| Db | 871 | YTVYRDGTKIKEGL-TETTFEEDGVATGNHEYCVKVTAGVSP-----KECVNVTV-- | 921 |
| QY | 233 | DHGDLNTPTVPMVLVAGAKFEKALPWLTKWAKQGYLDVHYTDEAEVGTNINAKFIH | 292 |
| Db | 922 | --DPVQNFQVNLTVSAVGQKVTLK----WDAPNG---TFNPNPNPNPGTTLS----- | 966 |


```

destruction and in bacterial host defense mechanisms. Activates
complement C3 and C5 (By similarity).
-!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
immunoglobulins, with a preference for Arg in P1, and hydrophobic
residues in P2 and P3.
-!- SIMILARITY: Belongs to peptidase family C25.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U85038; AAB41892.1; -.
EMBL; AS017173; AAO65700.1; -.
PDB; 1CVR; 01-MAR-00.
MEROPS; C25.003; -.
TIGR; PG0506; -.
InterPro; IPR007110; Ig-like.
InterPro; IPR001769; Peptidase_C25.
InterPro; IPR005536; Peptidase_C25_C.
Pfam; PF01364; Peptidase_C25; 1.
Pfam; PF03785; Peptidase_C25_C; 1.
Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen;
3D-structure; Complete proteome.
SIGNAL 1 24 BY SIMILARITY.
PROPEP 25 229
CHAIN 230 736 GINGIPAIN R2.
ACT_SITE 440 440 PROTON DONOR.
ACT_SITE 473 473 NUCLEOPHILE.
CONFLICT 58 58 G -> D (IN REF. 1).
CONFLICT 246 246 P -> A (IN REF. 1).
CONFLICT 251 251 E -> G (IN REF. 1).
CONFLICT 254 254 E -> K (IN REF. 1).
CONFLICT 398 398 I -> V (IN REF. 1).
CONFLICT 435 435 A -> V (IN REF. 1).
CONFLICT 480 482 YNV -> FSM (IN REF. 1).
CONFLICT 510 510 N -> D (IN REF. 1).
CONFLICT 512 512 S -> Y (IN REF. 1).
CONFLICT 515 515 S -> P (IN REF. 1).
CONFLICT 560 560 K -> N (IN REF. 1).
CONFLICT 582 582 K -> E (IN REF. 1).
TURN 237 238
TURN 240 245
STRAND 247 252
HELIX 254 262
TURN 263 264
TURN 266 271
STRAND 272 275
HELIX 281 293
TURN 295 296
STRAND 299 305
TURN 308 310
STRAND 314 316
TURN 317 318
TURN 319 321
STRAND 323 326
STRAND 338 342
STRAND 347 362
TURN 366 367
TURN 369 372
TURN 373 378
STRAND 382 382
TURN 384 385
STRAND 387 389
STRAND 391 391
HELIX 392 406
TURN 407 407
STRAND 410 415
STRAND 421 430
STRAND 434 438

```



```

Db 542 VEEMVDTPPTVTGGNIVRTVELGQSLRSVITYTEPTATDNSGEANLVSRTAQCDLFP 601
Qy 1143 IDLDEVEI-----KANGKRADETER-----PESTHGEA-----EW 1175
Db 602 VGTIVVEYVQDPAGNEGQGTIVTIVADITPPSVNCPNVVQVEVLTSSAPVFFTEP 661
Qy 1176 TTIDADGQGWLCSSGOLDWLTAGGNSVVSFWNGMALNPONYLISKDVGTATVK 1235
Db 662 TAFDISG-----ANLVTRTAAPG-----DSFPV-----GTSVS 691
Qy 1236 YYAVNDGPPG--HYAVMISKTGNAGDFTVVFEETPNKGARFGLSTEANGAKPQ 1293
Db 692 YIFSNNGNEAPCFTTIVISAVDT-----PPTVNCINN-----726
Qy 1294 SWIERTVLDPAKTYAFRHYNCSDNLNILLDDIQFTMGGSPTPTDVTYTVYRDGKIK 1353
Db 727 ---VAQTQVLGNGSTQVFTEPTAFDIS-----GQTSLVTRTSAGDSFP 768
Qy 1354 EGLTETFEEDGATGNHEVCVEVKTAGVSPKCVDTVTVNSTQPN-----VQNLTAEQ 1408
Db 769 VGTTSVTYIFTDASGNAQPCT-----FNVINAVDTTPTVNCANNIAQTI 815
Qy 1409 APNSMDAILKNAPASKRAEVLNEDFENGIPASWKTIADGDNWTTTPPPGSSFAGH 1468
Db 816 ELGSTAVNVAEPSA-----TDSGTAFVLVSRSSPGDSFPFGS 855
Qy 1469 NSAIQVSSASHINFEPP-----QNPDNVLVTPELSLPGGGLTFWVC 1510
Db 856 TTVTVIFSDQSGNEAPPCVFTTIGTVDTMMPFTSCPNIVQSVELGVP--GTWISW-- 911
Qy 1511 AQDANYASEHVAVYASSTGNDSANFANLLEVLTAQVTAPEAIRGTAQGTWYQKT 1570
Db 912 -----TTPTANDAAGIA-----SIVSNLQPGSFF--TV 937
Qy 1571 QLPAGTKVAFRHFECTDFEMLNLDVVITSGNAPSVTYTYRNTQTASGY--TETTYRD 1629
Db 938 GESATVTVADNSGLTD-----NSCSFTTVFAVDTPPPSVVCTNNVQFT 983
Qy 1630 PDLATGFTYGVKVVYPNGESAIEPATLNITSLADVTAQKPYTLTVVGKTTIVTCQGEAM 1689
Db 984 VELGTN-----PQVFEYTEPTASDISGQANLVSRTNV-----PDGSPFVG--TSVT 1030
Qy 1690 IYDNGRRLAQRNVTVTAQ 1710
Db 1031 FADNSGNTADACSEFTITVTAQ 1051

RESULT 7
WAPA_BACSU STANDARD; PRT: 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G OR BSU39230.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / BGSC1A1;
RC MEDLINE=95219088; PubMed=7704263;
RX

```

```

RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / BGSC1A1;
RA MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacXY region."
RL Microbiology 142:3113-3123 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.D., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrif P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpatra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L05634; AAA22883.1; -
CC EMBL; D31856; BAA0656.1; -
CC EMBL; D29985; BAA06260.1; -
CC EMBL; D83026; BAA11683.1; -
CC EMBL; Z99124; CAB15959.1; -
CC
CC DR
CC DR
CC DR
CC DR

```


DR PIR; S32920; S32920.
 DR Subtilisin; BG10797; wapa.
 DR InterPro; IPR003305; CBM_CenC.
 DR InterPro; IPR006530; YD_
 DR Pfam; PF02018; CBM_4_9; 1.
 DR Pfam; PF05593; Rhs_repeat; 14.
 DR TIGRFAMs; TIGR01643; YD_repeat_2x; 17.
 KW Cell wall; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 28
 FT CHAIN 29 2334
 FT DOMAIN 504 869
 FT REPEAT 504 605
 FT REPEAT 636 736
 FT REPEAT 769 869
 FT DOMAIN 1021 2139
 FT REPEAT 1021 1040
 FT REPEAT 1042 1061
 FT REPEAT 1063 1082
 FT REPEAT 1083 1102
 FT REPEAT 1109 1128
 FT REPEAT 1129 1148
 FT REPEAT 1150 1169
 FT REPEAT 1174 1193
 FT REPEAT 1199 1218
 FT REPEAT 1219 1238
 FT REPEAT 1646 1665
 FT REPEAT 1667 1686
 FT REPEAT 1690 1709
 FT REPEAT 1711 1730
 FT REPEAT 1732 1751
 FT REPEAT 1753 1772
 FT REPEAT 1795 1814
 FT REPEAT 1820 1839
 FT REPEAT 1840 1859
 FT REPEAT 1861 1880
 FT REPEAT 1887 1906
 FT REPEAT 1908 1927
 FT REPEAT 1929 1948
 FT REPEAT 1969 1982
 FT REPEAT 1983 2002
 FT REPEAT 2008 2027
 FT REPEAT 2028 2047
 FT REPEAT 2051 2070
 FT REPEAT 2071 2090
 FT REPEAT 2093 2112
 FT REPEAT 2120 2139
 FT SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;
 Query Match 2.6%; Score 242.5; DB 1; Length 2334;
 Best Local Similarity 19.4%; Pred. No. 4.9e-05;
 Matches 424; Conservative 265; Mismatches 750; Indels 743; Gaps 118;
 QY 14 GVGLYAQSAAK-----KLDAPTRTRCTNNKFKQFD-----A 45
 DB 88 GGVYKQEVLDPIHPTKPTDADWEDISPCLKESTSKQVETENAILNSDFQKQMKNGLYA 147
 QY 46 SFSNFVELT--KVETKGTFAVSIPGAFPTCEVGSPEVPAVKLIAPVPGATPVVRVK 103
 DB 148 TFEHNDHKVYSLAEAKGNKTSLTPKTSADYKTDSEI-----VYPDVFNIDLIQ 199
 QY 104 SFEQVYSLNQYSEKIMPHQ-----PMSKSDDPKQV----- 136
 DB 200 TET-----FENIKEDLVHQYNGYNTFTFLQKTDLQAKEQEDGSDFDSEKGVFVSVP 254
 QY 137 -PFVYN-----AAVARGFVGQELTQVEMLTGRGV-----RIALTINPVQ 178
 DB 255 KPFMTSDSKLDELSEGEVERSDKYSKLEKNEE-GYLLHLTADENWLKDPERVVPVSIDP-- 311
 QY 179 YDWMANQLKVRNRIEVEFQGADEVATORLYDASFSVP-FETAYKQLNRDVIYTDHGLD 237
 DB 312 ----STLSVSSDFTVMSAIPPTTNYASSQKQDANLKAIVLKTGY---YDKTTGTNYAFM 364

QY 238 -YNT--PVRMLVAGAKFEALPWLTKAQKGFYLDV---HYTDEAV--GTTNAGSI-- 287
 DB 365 KFNLLKPIQMTVTYKATLTKTYVAHSYGTGATGLWLDTVNSNY-DNAKVTWNTKPSAKNI 423
 QY 288 -KAFIHKYNDGLAASAAAPVFLALVGDVTVISGEKGGKTKKVTDLVYSADVDFPPEMYT 346
 DB 424 GKADVHK---GQWAS-----YDVAAVKS-----WNSGANYGFKLHT 459
 QY 347 FRMSASSPELTNIIDKVLMEKATWPKSYLEKVLIIAGADYSWNSQVQPTIKYGMQY 406
 DB 459 --NGNGKEYWKKLI-----SSANGANKPIEVTVYIPKGN-----TPTIK---AY 498
 QY 407 YNQEHGYTDV-----YNYLKAPYTSC-YSHLNTG-----VSFA 439
 DB 499 HNGDSTGYFISWKKVEGAGKYKWIYNGKBYQAIASAGNVTWSKGGKIWPTSAETASK 558
 QY 440 NYTAH-----GSETAW-ADPLLTSQLKALTNKDKYFLAI-----GNCCITIAQDFVQP 487
 DB 559 RYKHLHDGKDGAEALDPSVYKNSGGSYATSKN-YWIGVSAIPDOEGAMSAKPAKVP 617
 QY 488 CFGEVITRVKKGAYAYIGSSPNSYMGEDYYWVGANAV-FGVO-----PTFEGTSMGSD 542
 DB 618 NVG-----KAQAPSAKGYNNGNATGYF--DLSKAVSGATGYKVQVFNKGKGFETLDLGN-- 669
 QY 543 ATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHI GAHYWEAYHVL-GDGSVMPYRAMPX 601
 DB 670 ----QTSWTTKGGKIWPTS-AETKAGY-----ALHLKDGSGAEIPINPGP- 710
 QY 602 TNYTTLPASLPQNASYSIQASAGSYVAISKDGLVYGTGVANASGVAT-----VSMTKQI 656
 DB 711 --TYK-NAGDGAKNYSFK-----IIAYNKO-----GEALASPAATPALPIAPKPV 756
 QY 657 TENGNDVIVITRNSYLPVI-----KOIQVGEPSYPQVSNLTA-----TTQGVKVTLK 704
 DB 757 TGYLYTNTKSSQTYGNLIWEKVQNAKGYKVIYNGKEYQSPDVGDADEWTTQNKNI--- 813
 QY 705 WEAPSAKAEAGSREVKRICDGLFVTIEPA--NDVRANEAKVVLADNVDGNTGYQFLL 761
 DB 814 WPTSEIKA-GSYKLTGDKGGEALDPSVYNNANGN-----YKGGKNSFTL 861
 QY 762 DAHNTFGSVIPAT-----GLPFTGT-----ASSNLYSANFEYLV- 797
 DB 862 VA-YDANGETIPTAPENPTFHEGAELGTTEEYWSIIDPSQLNGATGVIVNEEDLSID 920
 QY 798 -----ANADPV-----TTOMIIVTQGE 816
 DB 921 GRGPGGLSRTYNSLSDSDHLFGQGYADAETSIVSTGGGAMYIDEDATHTFRFKKADGT 980
 QY 817 VVIPGVY-----DYCI-----TNP--BPASGKMIAGDGNQPA--RYDD----- 853
 DB 981 YQPTGVVLELTETADQFILTKQDQTNAYFNKKGKLOKVVDGHNNAVTYVYNDKNQLTA 1040
 QY 854 FTFEAGKYYTFT-----MRRAGMGDGTDMVEDDS----- 883
 DB 1041 ITDASGRKLTFTYDENGHVTSITGPKNKVTVSYENDLLKKVTDGTVTSYDSEGL 1100
 QY 884 -----PASYTTVYR-DGTLKEGLTATTFEEDGVAAGNHHEYCEVEKVTAGVSKV 933
 DB 1101 VKQYSANSTEAKPVTEYQYSGHRLKAINAK--KETVY---YSYDADKKTLMLTQPNG 1154
 QY 934 CKDVTVEGSNEFA-PVQMLTGSVGVQKVT--LKWDPAGTNPENPNPNPCT-TLSSEF 989
 DB 1155 RK--VOGYNEAGNPQI-QVIDDAEGLKITWTKYEGNNVVEDVDN---DVGTGKATESY 1208
 QY 990 ENGIPASWKTIDADGDGHGWKPGNAPGIAGNSNGCVYSSESFGLGGIGLVLPDNYLITPA 1049
 DB 1209 Q-----YDKQGNVTSVKDAYGTETETETKNNNDVTMKMDTEGNVTDIAD-----G 1253
 QY 1050 LDLPNGGKLTWVCADQADANYASEHVAIVAS-----STGNDASNFTNALLEETITAK---- 1100
 DB 1254 LDA-----VSETDQSGKSSAAVYDKYGNQIQSSKDLASATNLIKDGSEFAQKSGW 1304
 QY 1101 -----GVRSPKAIRIGT-----WRQKTVDLPAKTYVAF 1132

Db 1305 NLTASKDRRKSVIADKSGVLGSKALEVLSQSTSGAGTDHGYSSATQIVLEPNTITYLS 1364
Qy 1133 RHFOQSTDM-----FYDLDDEV-----EIKANKRADFTF-----TFSSSTHGEAPA 1173
Db 1365 GKIK-TDLAKSAYFNIDLRDKDQKRIQWINEYSALAGKNDWTKRQITFTTPANAGRAV 1423
Qy 1174 EWTIT-DAODGOGW-LCLSSGQL-DMLTAHGGSNVVSFWSMGMLNPDN----1221
Db 1424 VYMEVDHKDKGKAWPDEVQLEKEVSSNPVQNSFTSATENWNVSGASVDSERG 1483
Qy 1222-----YLSKDVGTGATKY-----1236
Db 1484 NDDVSLKAARTSASQAGSVTKQTVVLGQSANDKPVYLTLTGMSKASSVKFTDEKDYSLQA 1543
Qy 1237--YYA-----VNDGFP-----DHVAVMIKT-CTNAGDFTVVEEFTPNGINKGARFG 1282
Db 1544 NVYADSGTGYNAKFPSPGTQEWNAAVVVPKPKVNDKVSILFKSATG-----1594
Qy 1283 LSTEANGAKPOSVMIE--RTVDLPAGTKYVAFRHYNCSDLVYLLDDIQTWGGSPPTPD 1340
Db 1595-----TWPFDDIRLIEGSLTKST---YD-SNGNVVTKEDL---GYATSTD 1635
Qy 1341 Y-----TYT-----VYRDGTKI-----KEG-LTETTFEEDGVA 1367
Db 1636 YDETKKTSETDAKGETTYTYDQADQLTNMTLSNGTSLHSYDKEGNEVSKTIRAGADQ 1695
Qy 1368 TGNHEYCEVVEKYTAGVSKKCVDTVNSTOENPQNLTAEOAPNSMDAILKKNAPASKRA 1427
Db 1696 TYKFEYDVGKLVKTTDPLG-----NLASEYDANSLTKTISPNGNEVLSYDG-----T 1746
Qy 1428 EVLNEDFENGIPASWKTTIDADGDNNTTTPPGGSGFAGNSAICVSSASHINPEGPQN 1487
Db 1747 DRVKSXSYNGTEKYFTYDKNG-----NETSVNKE-QN 1779
Qy 1488 P-----DNYLVTELSLPGGGTLTFWYCAODANVASHYAVASSTGNDASNFANALLE 1541
Db 1780 TTKRTFKNRLTELTDRGGSQT--WTYPDSDKLTKFTSWIHGDKQ--TNQFTYKLD 1835
Qy 1542 EVLAKTVVTAPEAIRGTRAQGTWYQKVQLP-----AGTKY-----VAFRHGCT 1587
Db 1836 QMIEMKD-----STSSYSFDYDENGNTVQITNGGGSFTSYDERNLVSSLHIGDK 1886
Qy 1588 DFFWINLDDVVITSGNAPSYYTYIRNNTQI---ASGVTEYTRDPLATGFYTYG-----1640
Db 1887-----NGGDIL-----TESYEYDANGRTTINSASGKVQ-----YBYGKLNQ 1924
Qy 1641 -VKVVPNGESAIEATLNLITSLADVTAQKPYTLTVVGKTTITVTCQ-----GEAMIYDMN 1694
Db 1925 LVKETHEDG-FVIEYTYDVGNGRKTVTITIKDGSSTVNASFNIMNQLTKVNDESISYDKN 1983
Qy 1695 GRRLAAGRNVTVYTAQGHYAV 1716
Db 1984 GNRSDGKFTYTWDADNLTAV 2005

RESULT 8

ID YD96 METJA STANDARD; PRT; 2894 AA.
AC Q58791;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein MJ1396.
GN MJ1396.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;

Query Match 2.4%; Score 224; DB 1; Length 2894;
Best Local Similarity 18.2%; Pred.No.0.0007;
Matches 312; Conservative 213; Mismatches 562; Indels 626; Gaps 90;

Qy 214 FSPVFETAYKQLENRDVYTDHGLXNTPVRMLVVG---AKFKEALKPWLTKAOKGFYL 270
Db 792 WTSWSDIRYDNLNDSY---EGD-----AGTALAWTKDSLKPGEINWVPIWGL 838
Qy 271 DVHYTD---EABVGTT---NASIKAFIHKKYNDGLAASAPVF---LALVGDTDVISG 319
Db 839 GYNYTDMNEINMGLSQLYDVTGKSIDYPNNGDSFNPNGIPIIYINSTALYGLVAYNL 898
Qy 320 EKGKTKVKTDLIYSAVDGDF-----PEMYTPRMSASS-PEELTNIDK 363
Db 899 N-----VSINITQINGTYIYTNSTLNLISVPYEEKLVSPVNIISNPNYGVNITIK 950
Qy 364 V-LMYEKATWPKDKYLEKVLLAGADYSNNSVQGPQTKYGMQYVYQHEGYTDVN---419
Db 951 TNLNDQNTSD-----EKSIITVTSFVQPNQVQKTNVGBEIFYN-----ITLYNFGV 1001

QY 420 -----YLKAPYT-----GCYSHLNTGVSFAN-----YTAHGS- 446
 Db 1002 GGFEDINITLTKGTTTKIYNNSILIAEDANGDGDWYINPNYDLNSNLPDIIVPTGEI 1061
 QY 447 ETAWADPLLITISQKALTNDKDYFLATGNCCTIAQFDY-----VQPCF-----GEVIT 494
 Db 1062 NLTVSKTIPSTAPLGEIDTTTLTKFVNINNPISIFGKTTFTQSTPYPPSVQKTFYLRHGDTLR 1121
 QY 495 RVKEK-----CAYAYIGSSPNYSWGE-----DYWMSVG-----ANAVGVQ----- 530
 Db 1122 TLNTSIPPTINNYTINSLSIAHQPRFADNFTVVKPILLYINDPNVIFTEMHKI 1181
 QY 531 -----PTEGTSWGS-YDATFLEDS-----YN-TVNSI-----MWAGNLAATH 566
 Db 1182 VVSLMATNGIDSFILGSDVEYLYLDDTIKSYIFNLDSIITPKNYVVLVRVENQISSN 1241
 QY 567 AGNI-----GNTHIGHAYWEAYHVLGSGSNVPRAMPKTYTILPASLPONQASYS 619
 Db 1242 SINIYHNSYTPSNIT-LNTTIVVNVNIFSDKNVY-----LENENV-TIFANITDPIGSYD 1295
 QY 620 IQASAGSVVAISKDGVLXGTGVANAGVATVMTKQITENG-----NYDVVITRS-NY 671
 Db 1296 I--SGANITYVPNGSVI-----INSMLLQEIIDKXSPSLWKLYNYSFSLPESGKY 1344
 QY 672 LPVIKQIQVGEPSYPQVSNLTATTQOQKVTLKWEAPSAKAEGSREVXKRGIDGLFVTI- 730
 Db 1345 LITITGIE-----SNGVSKXNVTIYCGVEIQGVKED-----FCTLG 1382
 QY 731 -EPANDVRANEAKVLAADNVGNTGYOFLLDADHNTFGSVIPATGPLFTGTASSNLYS 789
 Db 1383 KESSEDKGIYGVNVSLEDS-----NNDGIPDIDG-----TIVNSTTDFGHYSFLVYN 1432
 QY 790 ANPEYLVPANADPVVTTQNIIVTQGEVWIPGGVYDVICITNPEPASGKMWIA-CDGQNQ 848
 Db 1433 SSKTYFVVVNSRTVTRGLNPQSKNDIWAETIYQTVYT-----PINSQMIANGASIFP 1489
 QY 849 ARY-----DDFTFEAGKYKTYTMRPAGMGDGDMEVE-----DDSPASYTYTVYRDGT 896
 Db 1490 DKLLATDDYD-EAGSVWYI--KPNLSE--DLVVEFYVLGDPDGDAGGITFTLQSLGT 1544
 QY 897 KIKGLTATTFEEDGVAAGNHEXCVVKYTAGVSPKVKDYTVVEGSEFAPVQNLTSVV 956
 Db 1545 -----NELGGTGGDLGY-----GGISPSVAVEYDT-WLNDP----- 1574
 QY 957 GQKVTLKWDAP-----NGTPNPN-----PNPENPGTTLSSSEFENGIPASW--- 997
 Db 1575 -----DAPATTDHIAIDVNGNINTYNSLTPTPNPDYDLGNVEDGREHLIKIWNAT 1626
 QY 998 -KITDADGDGH---GWKPGNAPGIAGYNSGCVYSSESFGLGG-----IGVLTDPNY 1044
 Db 1627 TKTLQVYFDGNLSLTWN-----KDTIIGNSAYFGTGGTGAKNLQVVKPIYVKNQDGY 1682
 QY 1045 LITPALDL-----PNGGKLTFWCAQDANYASEHYAVYASGTGNDASFTWALLEET 1096
 Db 1683 IINPTGVVEMFGGRDPN---EEDNW---EDGKY--EHYCL-INLNSYSGKNITFGSFDV 1734
 QY 1097 IT-----AKGVSPKAIRIGIOCTWRKTVLDPAGTKYVAFRHFSTDMFYIDLD 1146
 Db 1735 ITNKSNGQGSFQFIKNAALYKDSYFRIPIIDAKNGNHYI-----YTSGNKI---LD 1787
 QY 1147 EVEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGQGWCLSSGQDLWLTAGHSNV 1206
 Db 1788 NLAT-VNWS-----TQINGTII 1803
 QY 1207 VSSFSW--NGMALNPDNLYLSKDYTGATKVKYXYAVNDGPGDHYAVMISKTGTNAGDFT 1264
 Db 1804 LSGLOWTANGNA-----YI-----NNSNLT 1824
 QY 1265 VVEBETENGKKGARFGLSTANGAKPQSVWIERTVDLPAGTKYVAFRHFSCDLNLYL 1324
 Db 1825 LIL--TPDDYNQKG-----SW-----YKPNVLSDEL 1850
 QY 1325 LDDIQFTMGSGPTPD-YTYTVYRBDGTKIKEGLTETTTTTEEDGVATGNHCVKKTAGV 1383

Db 1851 VVEFYAVLGNPDGADGITFTLQSLGT-----NELGGTGGDLGY-----GGI 1892
 QY 1384 SPKCVDTVTNSQF-----NPVQMLTAEQAPNSMDA-----IL 1417
 Db 1893 SPVAVVEDTWLNDPDSFATTDHIAIDVDGNLNTHTYNSLTSTPNPYDLGNVEDGREHLI 1952
 QY 1418 K--WNPASAKRAEVLNEDFENGIPASWKTIDADGCGNNWTTTPPPGSSFAGHNSAICVS 1475
 Db 1953 KIVWNA-----TTKTLQVYFDGNLALTW-----NKDITQIIGNSTYFGTGG--TG 1996
 QY 1476 SASHINPEGP---QNPDNLYVTPELS-----LPGGG 1503
 Db 1997 GAKNLQVVKTYVKNQGNVNLNLEISEIPNPIIDNVGADTYIGNIFFENVSVGILGNETGLN 2056
 QY 1504 TLFFWVCA-----ODANYASEHYAVYASS--TGNDASNFANALLEEVLTAKTVV 1550
 Db 2057 NLTLKSGIYKILNAGVKLVYDWSLQNPVLYDLNLTINASGGYGISMLNKI----- 2109
 QY 1551 TAPEAIRGTQAQGTW---YQKTVQLPAGTKYVAFRHFPGCTDFFWIN-----LDDVV 1598
 Db 2110 -----WMLYNSQISLKNQ-----VGIYWANWAGFGNITTYNIT 2144
 QY 1599 ITSGNAPSITYT---IYRNNTQIASGVTTETTY 1627
 Db 2145 ISSCQGLVLYKQNGIKLINSQIKNSVYEGVY 2177
 RESULT 9
 IGAL HAEIN STANDARD; PRT; 1541 AA.
 ID IGAL HAEIN AC P42782;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Immunoglobulin A1 protease precursor (RC 3.4.21.72) (IGAL protease).
 GN IGA.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HK368 / Serotype B;
 RX MEDLINE=89379374; Pubmed=2506130;
 RA Poulsen K., Reinholdt J., Hjoorth J.P., Thøgersen H.C., Kilian M.;
 RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
 of Haemophilus influenzae serotype b.";
 RL Infect. Immun. 57:3097-3105(1989).
 RN [2]
 RP MUTAGENESIS OF SER-288.
 RC STRAIN=HK368 / Serotype B;
 RX MEDLINE=92234949; Pubmed=1373717;
 RA Poulsen K., Reinholdt J., Kilian M.;
 RT "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases.";
 RL J. Bacteriol. 174:2913-2921(1992).
 CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
 CC certain Pro--Xaa bonds in the hinge region. No small molecule
 CC substrates are known.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: Belongs to peptidase family S6.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X64357; CAA45708.1; -;
 EMBL; M87492; AAA24969.1; -;
 PIR; A37023; A37023.
 MEROPS; S06.001; -;
 InterPro; IPR006315; Autotransport.
 InterPro; IPR005546; Autotransporter.
 InterPro; IPR000710; Peptidase_S6.
 InterPro; IPR004899; Pertactin.
 Pfam; PF03797; Autotransporter; 1.
 Pfam; PF02395; IGAL1; 1.
 Pfam; PF03212; Pertactin; 1.
 PRINTS; PRO0921; IGASERPTASE.
 TIGRFAMs; TIGR01414; autotrans_bar1; 1.
 Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
 SIGNAL 1 25
 CHAIN 26 1008
 PROPEP 1009 1541
 ACT SITE 288 288
 MUTAGEN 288 288
 S->T: LOSS OF ACTIVITY.
 SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

Query Match 2.4%; Score 223; DB 1; Length 1541;
 Best Local Similarity 19.3%; Pred. No. 0.00031;
 Matches 324; Conservative 203; Mismatches 589; Indels 566; Gaps 87;

271 DVHY-----TDEAE-----VGTNASIKAFIHKYNDGLAAAPVFLALVGDTDVISEK 321
 31 DVDI QIPRFAENKGFSGVATNVLVK-----DKNNKDL-GTALPNGIPMI-DFSVDVDK 84
 322 -----GKK--TKKYTDLYSAVDG-----YFPEMYTFRM 349
 85 RIATLINPQYVGVKHSNGVSELHFGNLGNMNGNAKAHRDVSSENRFF----- 136

350 SASPEELTNIDKVLMEKAT-----MP--DKSYLEKVLIIAGADYS-----WNSQV 395
 137 SVERNEYPTKNGKTVTTEDQTKRRREDDYMPRLDKFVTEVAPTEASTASSDAGTYNDQ 196

396 GQPT--IKYCMQYVNOEHGYT-----DVNY--LKAPYTCYSHLN 433
 197 KYPAPVRIGSGQFYIKGDNYSLLNNHVEGNNLKLVDAYTYGIAGTYK--VNHN 254

434 TG-VSFANYTAHGETAWADPLLITTSQKALTNKDIFYLAICNCCITAFDYVOPCFGEV 492
 255 NGLIGFGN-----SKEHSDPKGIISQ--DPLTN---YAVLG-----DSGSLP 293

493 ITRVKEKAYAYIGSSP-----NSYGEDYVWSVGANAVFGVQTFEGTSMGSDATFL 546
 294 -VYDREKGLWFLGSDYDFWAGYNKKSQOE--WNI-----YKSQFT 330

547 EDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDSVMPYRAMPKTYT 606
 331 KDVLNKDS--AGSL-----IGSKT--DKSWSS-----NG-----KTSIT 361

607 LPASLPQNASYSIOASAGSYVAISKGLWYGTGVANASGVATVSMTKQITE-----N 659
 362 -----GGEKSLNVDLADG-----KDRPNHGSV--TEPGSGTLTNNDIQAGGLPFE 408

660 GNYDVITRSNVLPIVKIQVGEPSYQPVSNLTATTOGQKVTLKWEAPS-----AKKAG 715
 409 GDYEVKGSNTTWKAGSVVAE-----GKTVTWKHNPOYDRLAKIGK 453

716 SREVK-----RIGDGLFVTIEPAN-----DVRANEAKVVLADNVWGNTG 756
 454 TLIVEGTGDKSLKVGDTVILKQOTNGSGQHAFASGVIGSGSTLVINDKQVDENSI 513

757 YQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLYVPANADP-----VVTQN 808
 514 Y-----FG-----FRGRLDLNGSLTFDHI--RNIDDGARLVNHNMTNASN 553

809 IIVTQGEVWIPGGVYDYCIINPE---PASKMWIAGDGNQPARYDDFTPEAGKYTYT 865
 554 ITIGESLITDPNITPINDAPDEDNPAFR--IKDGGQLYNLNENYTYAURK----- 607
 866 MRRAGMGDTMEVDDSPASYTYTVYRDGKIK-----EGLTATTFEEDG 911
 608 -----GASTRSELPKNGSGESNENLYMKTSDAKRNVNMHNNERNMNGFVGEEG 661
 912 VAAGNHEXCVVEKYTAGVSPKVKCDVVEGSNE-----FAPVQNLTSVSGQKVTWKDA 966
 662 KNGN-----LNVTFGKSEQNRFLLTGGTINLGLDLYTEKGTL---F 700
 967 PNGTPNPNPNPNPGTT-----LSEFENGIPASW-----KTIDADGGHGWKPCNA 1014
 701 LSGRPTHARDIAGISSTKDPHAEANNEVVEDDWINRNPKATMMVTGNASYGRN- 759
 1015 PGIAGYNSGCVYSESGFGLGIGVLTDPNLYLITPALDLPNGGKLTFWYCAQDANYASEHY 1074
 760 --VANITSN--ITASNKAQVHIGYKTGT-----VCVR-----SDY 791
 1075 AVYASSIGNDASNETNALLBETITAKGVRSKATRGRIQGTWRQKTVDLDPAGTYVAFRH 1134
 792 TGYVTCT-----TDKLSDKALNS---FNPTNLRGVNLNLT---ESANFVLG-KANLFGT 837
 1135 FQSTDMFYIDLDEVEIKANGKRADFTETFFSSSTHGEAPAEWTTIDADGGQWCLSSGQ 1194
 838 IQSRGNSQVRLTE-----NSHWHLTGNSDVH-----QLDLANGH 871
 1195 LDWJTAGGGSNV-----VSSPFSWNGMALNPDNLYLSKDVGTAKTKVYIYAVN-DGFP 1246
 872 IHLNSADNSNNVTKYNTLTVNSLSNG-----SFYLTDLNSKQG 911
 1247 DHYAVMLSKGTNAGDFTVVFEEETPENGKGGARFGLSTEANGAKPOSWIERVDPAG 1306
 912 D--KVVTKSAT--GNFTLVADKTGPNHNLTLFDASKAQRDLHNLVSLVGVNTVDLGA- 966
 1307 TKY-----VAFRHYNCSDLNYLLDDIQTMGSSP----- 1336
 967 WKYKLVNNGRYDLYNPEVEKRNQTVDTNITFNENIQAQDVPSVPSNNEETARVDEAPVP 1026
 1337 -----TFTDYTYVRDG-----TKIKEGLTETTFEEDGVA-----TGNHEY 1374
 1027 PPAPATPSETTETVAENSKOESKTEVKEQOATETTAQNRREVAKEAKSNVKANTQNEVA 1086
 1375 -----VEKYTAGVSPKCVDTVNSTQFNPVQNLTAQAP--NSMDAILKWA 1421
 1087 QSGSETYETQTTETKETATVEKEEKAKVETEKQ--EVPKTVSQVSPQSQSETVQPOAE 1144
 1422 PASKRAEVLNEDFENGIPASWKITIDADGGDGNWTTTPPGSSPAGHNSAICVSSASHIN 1481
 1145 PARENDPTVN-----IKEPOSQNTTADTEQPAKETSSNVEQPVTESTTNTG 1192
 1482 FEGPQNDNLVITPELSLPGGTLTFWVCAQDANYASE-----HYAVASSTGND 1531
 1193 NSVVNDPEN--TTPATTQP-----TVNSENKPKNRRHRSVRSVPHNVEPATTSND 1243
 1532 ASNPA-----NALLEEVLTAQ--TVVTAPEALRGTRAQ-----GTWYQKV 1570
 1244 RSTVALCDLTSTWNAVLSDA--RAKAQFVALNVKAVSOHISQLEMMNEQYVWVSNT- 1301
 1571 QLPAGTKYVAFRHFGC-----TDFFWINLDDVITSGNAPSYTYTYIYRNNTQIASGVTE 1626
 1302 SMKNYSSSQYRRFRSSKSTQTLGW-----DQITSNVQLGGVFTYVRNSNFKATSKNT 1357
 1627 YROPDLATGYT-----YGVKVVPYNGESALET-----ATLNTISLADVTAKKYTTLTVG 1677
 1358 LAQNVFSKYIAONHWVILGIDLGKFPQSKLQTNHNAKFAHRTAQFGLTAGKAFNLGNFG 1417
 1678 KT 1679
 1418 IT 1419

| | | | | |
|------------|---|-----------|---|------|
| D | B | 840 | --SLNGITYSPAGG-----NFGTVPLLTNNKTGTQYVG | 871 |
| O | Y | 1063 | CAQDAN--YASHAYVAYSSTGDAS-----NFTNALLETITAKGVRS | 1104 |
| D | B | 872 | MVTDAAGNQTVDDQAIAFSSDKNALY-----NDISMKYLLRNISNVQV-DILDGGQ--- | 922 |
| O | Y | 1105 | PKAIRGRIGQWRQKTVDLDPAGTKVAFRHFQSDFMYIDLDEVEIKANKRADPFTETFE | 1164 |
| D | B | 923 | NKVTTLSNSTNRKKTYNAHSQQYI-YIHAPAWDGYYPDQRGNHKT---ADDSGYTR | 977 |
| O | Y | 1165 | SSTGEAPEAEMTTTTADGGCGWCLSSQLDMLTAHGGSNVVSSFSWGMLNPONLYI | 1224 |
| D | B | 978 | ISGVPEGDKRQVEDV-----PFKLDSKAPTVRHVALSAKTENG---KTQYLTL | 1023 |
| O | Y | 1225 | SK---DWTG--ATK-VKYVYAVNDGFPGDHVAVMISKTNAGDFTVVFEETP----- | 1271 |
| D | B | 1024 | AEKDDLUSGLDATKSVK--TAINENVLND--ATFTDAGTTADGYTKI--ETPLSDEQAQ | 1076 |
| O | Y | 1272 | ---NGINKGARFGLSTEANGAKPSQSVMIERTVDLPAGTKVAFRHYNCSDLNVILLDDI | 1328 |
| D | B | 1077 | ALNGNDN--SAELVLTDNASNATDQ----DASVQKPGSTSF-----DL | 1113 |
| O | Y | 1329 | QPTMGSGPTDYTYTVRGCTKIKEGLTEITFEEDGVAIGNHEYCVVEKYTAGVSPPKC | 1388 |
| D | B | 1114 | IIVGGGIPDKISSITTGYEANT---OGGGYTFF-----SGTPAAVDBGTYTDACQKKHD | 1164 |
| O | Y | 1389 | VDTVNS-----TCFNVPQN-----LTAEQAPNSM-----DALLKWNAPASKRAEVLN | 1431 |
| D | B | 1165 | LNTTYDAATNSFTASMVTNADYAAQVDLVADKHAHTQLKHFDTKVELTAPT----- | 1219 |
| O | Y | 1432 | EDEFNGI-PASWKTIIDADGGNNWTTPPPGSSFPAGHNSAICVSSASHINFEQPOND | 1490 |
| D | B | 1220 | LKENNSSDQTSEATIKVTGTVSADTKTVNVG-----DTVAALDAQHHFSVDVPVN--- | 1269 |
| O | Y | 1491 | YLVTPELSLPGGGTLTFVWCQAQANYASEHYAVVAS-----STGNDASN | 1534 |
| D | B | 1270 | -----YGMNIKVTATEDOENTTEOKTITSYDPMLEKNSTVFQGVTFGANEFN | 1320 |
| O | Y | 1535 | FANALLEVLTAKTVVTAPEAIRTRAQ-----GTWYQKTVOLPA | 1574 |
| D | B | 1321 | ATSARKFYDPKTGIATINGVKVHPHTTLQVDGKOPIPKDLLTFSTFLDLGLTGKQPFQVVU | 1380 |
| O | Y | 1575 | G--TKYVARPHFGCDEFPWINLDVVIT-----SGNAPSPTYIIYRNTOIASGVTTET--- | 1625 |
| D | B | 1381 | GDTQNKTFQ-----EALTFILOVAPRTLSDSSTDAPVYT---NNPNFQITGTATDNAQ | 1432 |
| O | Y | 1626 | ---TYRPDLATCFYTYGVKWVPNGESAIE-----TATNLNISLATDVTAQKP | 1671 |
| D | B | 1433 | YLSLSINGSSVASQYVDINIANSCKP-GHWALDOPVKLLBEGKNVLTAVVDSSEDNTTKMI | 1491 |
| O | Y | 1672 | TL-----TVVGKITTVTGOGEAMIYDMNGRRRLAAGRNTVVYTAQGG | 1712 |
| D | B | 1492 | TVYEPFKTLAAPFTVPSTTEPAKTVLTANSAA TG-ETVOYSADGG | 1537 |
| RESULT 11 | | | | |
| BIGA_SALTY | | | | |
| ID | BIGA_SALTY | STANDARD; | | |
| AC | R25927; P25928; Q9ACQ3; | PRT: | 1953 AA. | |
| DT | 01-MAY-1992 (Rel. 22, Created) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| DE | Putative surface-exposed virulence protein biga precursor. | | | |
| GN | BICA OR STM3478 | | | |
| OS | Salmonella typhimurium. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | | |
| OC | Enterobacteriaceae; Salmonella. | | | |
| OX | NCBI_TaxID=602; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=ATCC 14028; | | | |
| RA | Stojiljkovic I., Valentine P., Heffron F.; | | | |
| RI | "Salmonella typhimurium rhs homolog."; | | | |

Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2]

RL SEQUENCE FROM N.A.
RN STRAIN=LT2 / SGSC1412 / ATCC 700720;
RP MEDLINE=21534948; PubMed=11577609;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RT Nature 413:852-856(2001).
RN [3]
RL SEQUENCE OF 1-765 FROM N.A.
RP STRAIN=LT2;
RX MEDLINE=911100301; PubMed=1987123;
RC Wu J.Y., Siegel L.M., Kredich N.M.;
RA "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned *cysG* plasmid to overcome limiting *sir*oheme
RT cofactor.";
RT J. Bacteriol. 173:325-333(1991).
RN
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AF133696; AAD39458.1; -;
DR EMBL; AE008859; AAL22340.1; -;
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StyGene; SG10437; bigA.
DR Virulence; Repeat; Signal; Complete proteome.
KW
FT SIGNAL
FT CHAIN 1 27 POTENTIAL.
FT 28 1953 PUTATIVE SURFACE-EXPOSED VIRULENCE
FT
FT DOMAIN 101 252 PROTEIN BIGA.
FT REPEAT 101 103 15 X 11 AA TANDEM REPEATS.
FT REPEAT 104 113 1 (INCOMPLETE).
FT REPEAT 114 122 2 (INCOMPLETE).
FT REPEAT 123 133 3 (INCOMPLETE).
FT REPEAT 134 144 4.
FT REPEAT 145 155 5.
FT REPEAT 156 166 6.
FT REPEAT 167 177 7.
FT REPEAT 178 188 8.
FT REPEAT 189 199 9.
FT REPEAT 200 210 10.
FT REPEAT 211 221 11.
FT REPEAT 222 232 12.
FT REPEAT 233 243 13.
FT REPEAT 244 252 14.
FT REPEAT 252 267 15 (INCOMPLETE).
FT CONFLICT 207 207 D -> DRGDDVTPPD (IN REF. 1).
FT CONFLICT 514 514 A -> R (IN REF. 3).
FT CONFLICT 1698 1698 D -> N (IN REF. 1).
FT CONFLICT 1795 1798 QYLE -> ITLQ (IN REF. 1).
FT CONFLICT 1836 1837 SA -> T (IN REF. 1).
FT CONFLICT 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;
SQ
SQ SEQUENCE

Query Match 2.4%; Score 219.5; DB 1; Length 1953;
Best Local Similarity 20.2%; Pred. No. 0.00068;
Matches 266; Conservative 125; Mismatches 408; Indels 521; Gaps 70;
663 DVVITRSNVLPIVK-----QIQVGEPSPVQPVSNLTATQGGKVTLK--- 704

```

OX
NCBI_TaxID=602;
RN
[1]_
SEQUENCE FROM N.A.
RP
RC
STRAIN=ATCC 14028;
RA
Stojiljkovic I., Valentine P., Heffron F.;
RT
"Salmonella typhimurium rhs homolog.";

```

Db 251 DSVITFNGVTIDKDKTLTDFSKLDNGSVLEGAVWNYSEQDNQMQLTTADGKTLNVGT 310
QY 705 WEAPSAKKA--EGSREVKRIGDGLF-----VTEPANDVRANEAK----- 742
Db 311 WDVTDAANAIEGTQE-----NGLYWKYDSRGYLIADDDNTTVISGDDQAHNSDRGMDS 365
QY 743 -----VYLAADNVWMDNTGYQFLLDADHNTFGSVIPATGPLEFTGTASSNLYSANFEY-- 794
Db 366 GQDRTGVLISGDRVNTLTGSSVTDG-----ATGMVISGDGTTTNSHSTVDN 415
QY 795 -----LYPANADPVVITQNIIVTGGVWVPGVVDYCIITNPERASGKWITAGDGGNQPAP 850
Db 416 ATGALISGNG-----TTTNF-----AGDIAVSGGGTAIIII-----DGDN----- 449
QY 851 YDDFTFEAGKKYFTMRRAGWG-----DGTDMVEDEDDSPASVYTYVYRDGTIKKEGLTA 904
Db 450 -----ATIKNTGTSIDISGAGSTGTWIDGNARVNNDGDMTIT-----DGGTG 491
QY 905 TTPEEDGVAAGNHEYCVVEVKYTAGVSPKVKDVT---VEGSNEFA---PVQNLTGSSVGQ 958
Db 492 GHTTGDNVVIDN-----AGSTTVSGADATALLYIEGDNALVINEGNTQISSGGAVT 541
QY 959 KVTCLKWDAPNGTNPENPNPNTLSESFENGIPASWKTIADADGCHGWKPGNAPGIA 1018
Db 542 RI-----DGDG-----A 548
QY 1019 GYNSNGCVYSEFGLGGIGVLTPDNYLITPALDL--PNG--GKLTFWVCAQDANYASEHY 1074
Db 549 HTTNTGDIADVAGSAV--IINGDNGSLTQAGDLLVTDGAMLIITY----- 593
QY 1075 AVYASSTGNDASFTNALLEETTAKVRSKPAIRGRIQGTWQK-TVDLPA-GTKYVAF 1132
Db 594 -----GTGNEAKNTGNA-----TVRDADSVGFVAGEKNTFKNKGIDIVSLNGTGAUV- 641
QY 1133 RHFQSTDMFYDLD-EVEIKANGKRAEDFTETESSHGEAPAEWTTIDADGD-----G 1184
Db 642 -----SGDMSQVTLGDINNVIS---VQDSEGVFSATGVSVGSNVAVIDTGNVNI SADY 694
QY 1185 QGWLCLSSGOLDLTAHGGSNVYSSFSWNGMALNPONLYLSKDVGTAKTKYKYVAVNDGF 1244
Db 695 QDDLAAGAPLITGVVVGNGNTV---TLNG-ALNIDNDNIS-----ATGGQYLDVGLSV 745
QY 1245 PGHYAVMLSKTGNAGDFTVFEETP-----NGINKGARFGLSTPANGAKPQSVW 1296
Db 746 TGDNDNVEID-----GGINITHSEDPDGTSDADITGISVSG---NSTVTLNG----- 789
QY 1297 IERTVDLPAGTKVAFRHYNCSDLN---YILLDDIQFTWGGSPPTD-VTYT--VYRDG 1349
Db 790 -HSTID-----TNTVVGHHVVLARVNGGSLILGDDSVVDVNVSVIPTGYTYTNALLMADG 844
QY 1350 TKIKEGITETTFBEDGVATGNHEYCVVEVKYTAGVSPKKCVDTVNVSTQFNPQN----- 1403
Db 845 -----EG---TSIENKGDITSHGVVSV-IRADNGSEVSNSGDILVYATSSNSEDRAAITR 896
QY 1404 -----LTAEOAPNSMDAI-----LKW----- 1419
Db 897 ASGEGSAVHNKAGDITLISDQTFQSGGIEVYPLKWTHTFYAMNASDYGDVNVDEGAT 956
QY 1420 -----NAPAKRAEVLNE-----DFENGIPAS--WK----- 1443
Db 957 IHLQAGAGVYGTASRGKALNEGNIYLDGLVPTLDDENNITSTSYWQSSLYLTSSGMVAG 1016
QY 1444 TIDADGSG-----NN-----WTTTPPG----- 1461
Db 1017 STDADGDATAINTGNITVNNAGFGMALNGGTAINQGVITLTADGVTGQADELVGMAL 1076
QY 1462 -----GSSFAGHNSALCVSASHINEFGP--QNPDNYL-VTPE--- 1496
Db 1077 NGGVWINDTSGVINIDADYGOALFSDSSSYIINNGS--INLNGSPMDDTDSHMGGTPTDKI 1135
QY 1497 --LSLPGGG---TLT---FWVCAQADANYASEHYAVYVYASSTGNDASNFANALLEEVLTA 1547
Db 1136 WIQSLFGSGSDRTSDTGTGFTAGTLANVGTETLNGVDVNVGGLVNEAGASL----- 1188

QY 1548 TVVTAPEAIRGTRAQGTWYQKTYQLPAGTKYVAFRHFPGCTD-----FFWIML----- 1594
Db 1189 -----TVNG-----TINGGAN--ALANYGTILDADAISTWHSLEFNEADGSIT 1229
QY 1595 DDVITSGNAPSYYTITRNNTQIASGVTTETTYRDPDLATG-----PYTY 1639
Db 1230 TDLTLNGDVTFY-----NNGDFTGSIAGTSVQOEIVNTGDMTVADBGKSLVSGSYFY 1283
QY 1640 GVK-VYVPGESAI-----EYATLNTSLADVTAQKPYTLTVGKTIITVTCOGEAMIYDMNG 1695
Db 1284 NEEDATLTGSGSAVEGENTIINLTRANDSLQ-----VNSGITAT-NGYSAITTVNG 1336

RESULT 12
YS89 CAEEL
ID -YS89 CAEEL STANDARD; PRT; 3178 AA.
AC Q09624; Q09625; Q0969D4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -1- SIMILARITY: Contains 1 PLAT domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48544; CAB70192.1; -
DR EMBL; Z48582; CAB70192.1; JOINED.
DR EMBL; Z48582; CAB70201.1; -
DR EMBL; Z48544; CAB70201.1; JOINED.
DR WormPep; ZK945.9; CE25697.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR005821; Ion_trans.
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR InterPro; IPR003915; PKD_2.
DR InterPro; IPR000203; PKD_cys_rich.
DR InterPro; IPR008976; PLAT_LH2.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF01477; PLAT; 1.
DR PRINTS; PR01433; POLYCYSTIN2.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00308; LH2; 1.
DR PROSITE; PS00095; PLAT; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 51 73 POTENTIAL.
FT TRANSMEM 2139 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.

FT TRANSMEM 2483 2505 POTENTIAL.
 FT TRANSMEM 2567 2589 POTENTIAL.
 FT TRANSMEM 2836 2858 POTENTIAL.
 FT TRANSMEM 2939 2961 POTENTIAL.
 FT TRANSMEM 2976 2998 POTENTIAL.
 FT TRANSMEM 3038 3060 POTENTIAL.
 FT DOMAIN 266 1196 SER/THR-RICH.
 FT DOMAIN 1105 1241 GLY/SER-RICH.
 FT DOMAIN 2071 2120 GPS.
 FT DOMAIN 2182 2305 PLAT.
 SQ SEQUENCE 3178 AA; 344726 MW, F8239436D03666CD CRC64;

Query Match 2.4%; Score 218.5; DB 1; Length 3178;
 Best Local Similarity 19.8%; Pred. No. 0.0016;
 Matches 285; Conservative 185; Mismatches 525; Indels 443; Gaps 72;

QY 426 TGCYSHLTGVSFANYTAHGSSTAWADPLLTTSOLKALTNKDKYFLAIGNCCTIAQPDYV 485
 DB 501 TGPSSSSTPSTASSV--SSTASSTQSSTTQSSSTTTTKE-----TTTSSD-- 547
 QY 486 QPCFGEVITRYKEGAYAYIGSSPNSYMG-----DYVMSVGN-----AVFGVQPTFE 534
 DB 548 -----GTPDPFYFEXATTFTFYDSTVNLTLNSGLGIIGYQISIE 587
 QY 535 GTS--MGSYDAFTLEDSYNTVNSIMWAGNLAATHAGNIGNTHIG-AHYWEAVHVLGDG 591
 DB 588 CTSPTSNYSVTTKDGACFTKSVSM--PRLGCTYPAS-----TFVGPNGYTFRATMTTDDK 641
 QY 592 SVMYRAMPKNTVTLPASLPQOASYSIOASAGSYVAISKDGLVYCGVGNASGVATVS 651
 DB 642 KVV-----YTYANVYIQEVSSTTIES-----STSAVASSTSPSTPSSSTLS 685
 QY 652 MKOITENGNDVYVITRSNLPVLIQIQVGPSPYQPVSNLTATTOGOKVTLKWEAPSAK 711
 DB 686 -TSVWTEPSS--TRS-----SDSTTISAGSTTLTQSTTISE 719
 QY 712 KAEGSREVKRGDGLFVIEPANDVRANEAKVVLAAADVWGDNTGYQFLDAD----- 764
 DB 720 ESTTDSSTTISLSTSTSSPSSTTA--DSTSLSDVQ-----FDFILDSGLSWNETR 770
 QY 765 HNTFG-SVIPATGLFTCTASSNLSYANFELVLPANADPVVTTQNIIVTGOGEVVIQGV 823
 DB 771 HNEDSINIVLPNTAIFPERSQFECR-----NVSTFPFLIKESTCL-----N 815
 QY 824 YDCITNPEPAGSKWMIAGDGNOPARYDDFTFEBAG-KKYFTFMRAGMGDGTMEVEDD 882
 DB 816 YSNVTLNATYSS-----NIPQPIE-TFLVGIGTYEF---RINMTDLTTMQVVSH 861
 QY 883 SPASVTVYVREDGKIKIEGLTATTFEEDGVAAGNHCYCVKVTAGVSPKVKCDVTVEGS 942
 DB 862 ---IFTLVNVDSTSE-VTSIT-----STGSSSSSAISTTSGIB-----STSTLEAS 907
 QY 943 NEFAPVQNLTGSS-----VGQKVTILKADPNATPNPNPNPNPGTTLSEGFENGIPASWK 998
 DB 908 ITDASQDSSTSDSGTSDSTTI--DSNST-----PSTDSSGLSQTPSDSSASASMR 961
 QY 999 TIDADGCHGWKPNAGIAGYNGCVYSBESFGLGIGVLPDNYLITPALDLPN--GG 1056
 DB 962 TTTVDPPASTPTDFV-LENLTWNETVY-----YSENPFIYTP--IPNKEPG 1006
 QY 1057 KLFTEWCAQDANYASEHY-----AVVASSTGNDASFTNA-----LLE 1094
 DB 1007 ALTTAMTCQCNDSQFVLKESNCITFEKNGAYASVSFNPMTSFVPTGTYEFLIN 1066
 QY 1095 ETITAKGVSPKAIIRIGTQRTWRTKVTDLPAKTKYVAFRHFQSTDMFYDLIDEVEIKANG 1154
 DB 1067 VTNEASGESASHIF-----TMNVVLPITTTTPTTVSSSD-----DAGG 1106
 QY 1155 KRADFTTFSSSTHGEPAEWTTIDAGD-----CQGLCLSSGQLDMLTARGGSNVV 1207
 DB 1107 KTGGTGAT--GGTGGTGGGSAATLSTGDAVRSTTSGS-----SGQSSTGSGAGSGSTT 1159
 QY 1208 SPSFWNGMALPNPNYLISKDVTGATKYKYVAVNDGPFGBDHYAVMISKGTGNAGDFTVVF 1267

DB 1160 ASGSGS-----GSSGTGSD-----GVNSG-----KTTALNDGGTSGTAT--- 1195
 QY 1268 EETPNG-INKGARFGLSTEANGAKPQSVWIERVTLDPAGTKYVAFRHYNSDLNIIILLD 1326
 DB 1196 --TPGSHLDGGSTSGSGSDSGSGSVS-----TK----- 1223
 QY 1327 DIQTMGSPPTDYTYTVYRDGTKIKEGL-----TETTFEEDGVATGNHEYCV 1375
 DB 1224 ----SSGSDT-----SGSSDSGSGANGAFSAQAQPSRTTRTKRSSLAT----- 1262
 QY 1376 EVKYTAGVSPKKCVDTVTNSTQFNPQVNLTA-----EOPNSMDAILKW--NAPASK 1425
 DB 1263 -----VSPISAAEQALIDAQAQADVMNQLAGIMDGSASNNLSNTSSLLNQISLPAAD 1315
 QY 1426 RAEVLNEDFENGIPASWKTIDADGDN-----NWTTPPPGGSSFFAGHNSAIC 1473
 DB 1316 LVEVAQSLLSN-----TLKIPGVGNMSSVDVLKTLQDNIATT-----NSELADEMAKV 1363
 QY 1474 VSSASHINEFGPQNPNDVMTPELSLPGGTLTFWVC-----AQDANYA-----SEHY-- 1521
 DB 1364 ITKLANVMTSAQSLNSVLSLDLALAGSTVYTLGVSVSTSKOGTYAVIFGYVIASGYTL 1423
 QY 1522 -----AVYASS-----TGNDASNPANALLBEVLTAKTWVTAPEAIRGTAQGTWYQKTVQ 1571
 DB 1424 VSPRCTLSIYGSTIYLTGDTRASV-KQLDGDVTYADTMLAAAGIQGFATN---GRTVQ 1479
 QY 1572 -----LPAGTKYVAFRHFPGCTDFFWNLDDVVITSGNAPSYYTYTYRNTQIAS 1620
 DB 1480 VEQKIDDKRSLVSGNIMATMSGVDVQSGEYSYNDMYTANV---TY-----DNSTVGS 1532
 QY 1621 GVTEYTVRDPOLATGFTYGVYKVVYPNGESAIEPATL-----NITSLADVTAKPYTL 1673
 DB 1533 TSQKNT-----SFSFNIPSEVQYIILLISGTMKLSHTQNTVSRGLV---TA 1578
 QY 1674 TVVGTITVTCQGEAMIYDMGRRLAAGRTVTVYTAQGHYAVVWVVDGK--SYVEKL 1729
 DB 1579 SYGGVTVTITCT-----NCTGKFVEVDTDNAIFSYNADSTFTV-VASDGSASVTKL 1629

RESULT 13
 APU_THETU
 ID - APU THETU STANDARD; PRT; 1861 AA.
 AC P38536;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Amylopullulanase precursor (Alpha-amylase/pullulanase) (pullulanase type II) [includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].
 GN AMVB.
 OS Thermoanaerobacter thermosulfurogenes (Clostridium thermosulfurogenes).
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 OC Thermoanaerobacteriaceae; Thermoanaerobacterium.
 CX NCBI_TaxID=33950;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=DSM 3896 / EM1;
 RX MEDLINE=94252998; PubMed=8195085;
 RA Matuschek M., Burchardt G., Sahn K., Bahl H.;
 RT "Pullulanase of Thermoanaerobacterium thermosulfurogenes EM1 (Clostridium thermosulfurogenes): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface.";
 RL J. Bacteriol. 176:3295-3302(1994).
 CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic linkages in pullulan and in amylopectin and glycogen, and the alpha- and beta-limit dextrins of amylopectin and glycogen.
 CC -I- COFACTOR: Binds 1 calcium ion per subunit (By similarity).

CC --!- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN
 CC --!- PTM: Glycosylated.
 CC --!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
 CC --!- SIMILARITY: Contains 2 fibronectin type III domains.
 CC --!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M57692; AAB00841.1; -
 DR HSSP; Q08751; 1BVZ
 DR InterPro; IPR006589; Alp_aml_cat_sub.
 DR InterPro; IPR006048; Alpha_aml_C.
 DR InterPro; IPR006047; Alpha_aml_cat.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR004185; Glyco_hydro_13lg.
 DR InterPro; IPR004193; Glyco_hydro_13N.
 DR InterPro; IPR001115; SLH.
 DR Pfam; PF00128; alpha-amylase; 1.
 DR Pfam; PF02806; alpha-amylase_C; 1.
 DR Pfam; PF02903; alpha-amylase_N; 1.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF02922; isoamylase_N; 1.
 DR Pfam; PF00395; SLH; 3.
 DR SMART; SM00642; Amy; 1.
 DR SMART; SM00632; Amy_C; 1.
 DR SMART; SM00060; FN3; 2.
 DR PROSITE; PS01072; SLH DOMAIN; 3.
 KW Carbohydrate metabolism; Multifunctional enzyme; Hydrolase;
 KW Glycosidase; Calcium-binding; Repeat; Signal; Glycoprotein.
 FT SIGNAL 1 35 POTENTIAL.
 FT CHAIN 36 1861 AMYLOPULLULANASE.
 FT DOMAIN 928 1018 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 1157 1248 FIBRONECTIN TYPE-III 2.
 FT ACT_SITE 628 628 BY SIMILARITY.
 FT ACT_SITE 657 657 BY SIMILARITY.
 FT ACT_SITE 734 734 BY SIMILARITY.
 FT METAL 401 401 CALCIUM (BY SIMILARITY).
 FT METAL 403 403 CALCIUM (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 406 406 CALCIUM (BY SIMILARITY).
 FT METAL 407 407 CALCIUM (BY SIMILARITY).
 FT METAL 452 452 CALCIUM (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 453 453 CALCIUM (BY SIMILARITY).
 FT DOMAIN 1681 1739 SLH 1.
 FT DOMAIN 1740 1803 SLH 2.
 FT DOMAIN 1804 1861 SLH 3.
 FT CONFLICT 1734 1734 D -> E (IN REF. 1; AAB00841).
 SQ SEQUENCE 1861 AA; 206104 MW; 06C23070E453B574 CRC64;

Query Match 2.3%; Score 214; DB 1; Length 1861;
 Best Local Similarity 18.8%; Pred. No. 0.0013;
 Matches 332; Conservative 200; Mismatches 638; Indels 598; Gaps 87;
 QY 139 VYNAAYARKGVGQELTQVEMLTGMRGVRIALTINPVQDVANQLKVRNRIEVSF 198
 DB 178 VYSYTAHVPGDYQYKVT-----LG-----NTWDENYGANGVKDGSNIQINVT- 220
 QY 199 QGAEVATQRLYDA-----SFSYFETAYKQLENRDVYTD-HGDLNTPVRLMWAGA 250
 DB 221 ----NDADITFYDANHTNTWNTSYPLTGLDNNIYYDLKHTDHSFFRNFGAVKV--- 274
 QY 251 KFKEALKPWLTKAQ-KGFYLD---VHYTDE---AEVGTGTA-----SIK-A 289
 DB 275 -----DQIVTLRIQAKHDLSEARISYWDINKIRIELPMTRIGESPDGPNFEYWEIKLS 328

QY 290 FIHKK-----YNDGLAASAAPVFLALVGDTVISGEKGGKTKYKVTLDLYSAVDGYFP 342
 DB 329 FDPTRTWYFIILKDGTKT-----AYGNDQDLGGVGKAT-----DTNKDF-- 371
 QY 343 EMTFRMSASSPEELTNIIDKVMYKATMPDKSYLEKVLIIAGADYSWNSVQGPITKY 402
 DB 372 ELATYDKNEDTDPWMKGV---MYQ--IPDFRYNGD-----TSNDIAKTLRGNDFIE- 420
 QY 403 GMQYVYNOEHGYTDVINYLYKAP-YTCYSHLNTGVSFANYTAHSETAWADPLLTSQLK 461
 DB 421 -----FHNNDLDPNPNAGTGYTG-----DGIWNSDF--GDLK 455
 QY 462 ALTNKQKYLAIAGNCCT-----TAQPDYVQPCSG-----EVTIRVKEKG 500
 DB 456 GIDDKLDYKGLGVSYLYNPIFESPNHKYDADYTKIDEMFGTTQDFEKIMSDAHAKG 515
 QY 501 AYAYIGSSPNSYMGEDYY-----WSVCANAVFGVQPTFECTSGSYDATF 545
 DB 516 IKIILDGVENHTSDDSIYFNRYGKYPCLGAYQAWKEGNSLSYGDWYNTINSOGTYECWW 575
 QY 546 LEDSYNTVNSI-----MWA-----GNLAATHAGNIGNITHIGAH 579
 DB 576 GYDSLPIVKSLNGSEYNVTSMANFIINDENAIKYLNPDLNLDGADGWLVDVENEVAH 635
 QY 580 YYWEAYH-----VLGD--GSYMPY----- 596
 DB 536 DFWTHFRNAINTVKFPAPMAENWGDASLDLGDSPNSVWNYOFRNDIIDFLGQSPDDG 695
 QY 597 -----RAMPKINTYTLPAFLPQNASYSIQASAGSYVAISKDGLVYGTGVA 642
 DB 696 NGQHNPIDAAKLDQRLMSIVERYPLP-----AFYSTWNLGSHDTMR---ILTVPGYN 745
 QY 643 NASGVATVSMTKQITENGNDYDVVITRSNLPVKIQVGEPSY-----OPVSNLTATTQG 698
 DB 746 SADPNENSDAAKLABQKLKATILQMGY-PGMADIYYGDEAGVSGKDPDDRRTPFWGN 804
 QY 699 QKVTLKWEAPSAKKAEGSRVKRIGDLFVTTEPANDVRANEAKVVLADNVWGDNTGYQ 758
 DB 805 EUTTLQDFFKNTSSIRNNQVLTGKD--LETVAQNDVTAIGRIINGKD---AFGTSYP 859
 QY 759 -----FLDADHNTFGSIVPATGPTGTASSNLXANFEYLVPANADPVVTTQNIIVTG 813
 DB 860 DSAAIIVAINRSKDKQIAIDTTKFLRDGVTFKDLINNVSYSI-----S 903
 QY 814 QGEVVIPEGVDYCYTINPEPA-SGKMWIAGDGNQPARYDDTFEAGKKYTFMTMRAGMG 872
 DB 904 NGQIVT-----DVPAMSGVMLISDDG-----QDLTAPQAPSNVVVTSNGKV 945
 QY 873 DGTDMVEDDSDPASVYTYVRDGT-----KIKEGLTATTFEEDGVAAG-NHEYCVVEVKY 925
 DB 946 DLSWLQ-----SDGATGYNIYRSVVEGLYEKIASNVTTETTFEDANVTGLKTVYALSAID 1001
 QY 926 TAGVSPKVCQVTVESNEFAVPQNLTGSSVGQKVTWKDAP-----NGTPNPNP 975
 DB 1002 ELGNESGISANDAVAYPAVPIGWVGNLT--QVSDNHIIGVDKPTEDIYAEVWADGLTN-ST 1058
 QY 976 NENPNPGLTSESFENG-----IPASWKTI--DADGDGHGWKPGNAPGIAGNSNGCVYS 1028
 DB 1059 GGGPNMIAQLGYKYSGTYGVSYSVNSVNSVGVDSGFTWNAQYVGDIGNNQ---YK 1115
 QY 1029 ESFGLGGIGVLTDP-----NVLII-----TPALDLPNGGKLTFWVCAQD-----ANYASE- 1072
 DB 1116 ASF-----TPDKIGQWEIYLMRFSNQGDWITTSLSFYVVPDSDLIKETAPYLQNP 1167
 QY 1073 -----HYAVYASSTG--NDASNFTNALLE--ETITAKGVRSRPA 1107
 DB 1168 GTSESRVSLTNWPNSTDNVGIDYEIYRSDGGTFNKIATVSNVEVNYIDTSVINGV----- 1222
 QY 1108 IRGRLOGTWRQKTVLDLPAGTKVAFRHFQSDTDMFVIDLDEVEIKA--NGKADPTE----- 1161
 DB 1223 -----TYNTKVAVD-----LSFNRTES--NVVTIKPDVVPKIVFNVTVPDYTPDAVN 1269
 QY 1162 ---TPESSTHGEAPAEWTTTIDADGGQGLWCLSSG-QLDWLTAHGG----- 1203

```

Db 1270 LAGTFPNATWPSAQMTKID-NNTYSITLTLDEGTQTEYKARGSWDKVEKDEYGEFA 1328
Qy 1204 -----SNV-----VSSFSWNG-MA 1216
Db 1329 SNRKVTIVNQNNEMTINDTVWRDIPFIYSPSSNMTVDNSISTMEVKNGTYKGAQVT 1388
Qy 1217 LNPNY-----LISKQVT---GATKKYIYAVNDGFFGDHYAVMISKGTGNAGDFT--- 1264
Db 1389 INGSFVQDKNGVFTKDVSLNKGVNKIKIHVEPNDG-----SYVGNQDGRITELT 1438
Qy 1265 -----VVFETPTGINKGGRFGLSTFANGAKPSQWIERTVLDPAGTKYVAFRHVNC 1318
Db 1439 KOIIVIRQENNSGSGGNN-----NTSTSGSNSS----- 1470
Qy 1319 DLNVLDDIDIOFTMGSGSPPTDYTVVYRDGTKEIGLTTETFEEDGVATNGHEVCVEVK 1378
Db 1471 -----TSGSGTGSTSIISNIN-----TSNINVTIGVITKNGN-VITLT 1508
Qy 1379 YTAGVSPKCVVDVTVNSTQFNPVQNLTA-----EQAPNSMDAILKW-N 1420
Db 1509 LDAG-----KAKDLIVNSKDKKWFEDITIGBQQKVVQISKDILDTSAANGKDIVIKSDN 1564
Qy 1421 APASKRAEVLEND-FENGIPASWKTIDADGDN--NWTTPPPGSSPAGHNSAICVSSA 1477
Db 1565 ASIATKDALNQIQNGVNVVS-----IKDNGKPNVTNVSLSNVVDITISGISGNTVLAKP 1621
Qy 1478 SHI--NFEQGPONPNLY-----TPELSLPG-----GTLTFWCAQADANYASEHYAV 1523
Db 1622 VEVTINISKANDPRKAVVYVYNNPTNQWEYVGGKVDASSGIIIF-----NATHESQVAAF 1676
Qy 1524 YASSTGND-ASNFANALLEEVLTAKTVV 1550
Db 1677 EYDKTFNDIKDNWAKDVI-EVLASRHIV 1703

RESULT 14
YEEJ_ECOLI
ID_YEEJ_ECOLI STANDARD; PRT: 2358 AA.
AC P76347; P94750;
DT 28-FEB-2003 (Rel. 41, Created)
DE 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeeJ.
GN YEEJ OR B1978.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayaashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392 (1996).
CC -!- SIMILARITY: Contains 13 B12-1 domains.

```

```

CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000289; AAC75042.1; ALT_INIT.
DR EMBL; D90837; BAA15800.1; -.
DR EMBL; D90836; BAA15799.1; ALT_INIT.
DR EcoGene; EG13378; yeeJ.
DR InterPro; IPR003344; B12_1.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR003482; LysM.
DR Pfam; PF02369; B12_1; 13.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; B12_1; 13.
DR SMART; SM00257; LysM; 1.
DR SMART; SM00089; PKD; 6.
DR KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 931
FT DOMAIN 932 1033
FT DOMAIN 1042 1137
FT DOMAIN 1146 1237
FT DOMAIN 1246 1350
FT DOMAIN 1351 1448
FT DOMAIN 1449 1553
FT DOMAIN 1554 1655
FT DOMAIN 1661 1754
FT DOMAIN 1763 1853
FT DOMAIN 1855 1950
FT DOMAIN 1952 2053
FT DOMAIN 105 105
FT CONFLICT 105 105 S -> G (IN REF. 2).
SQ SEQUENCE 2358 AA; 248599 MW; 232249750BF631ED CRC64;

Query Match 2.3%; Score 211; DB 1; Length 2358;
Best Local Similarity 19.3%; Pred. No. 0.0026;
Matches 371; Conservative 212; Mismatches 700; Indels 642; Gaps 88;

Qy 27 DAPTRTCTCTNNFKQFDASF-NEVELTKVETKG-GT 63
Db 534 DSSVSLSTQTLNADSHSTATLFIADAGNPVGLVSLSTRHEGVQDITLSDWKNDGGS 593

Qy 64 FASVSIPIGAF-----PTGEVSPSEVPAVKLIIV-----PVG 95
Db 594 YIQILATGAMSGTLTLPQLNGVDAAKAPAVNIISSVSRTHSSIKIDKRYLSGNPIE 653

Qy 96 ATPVVR---VKSFTQVYSLNQVSEKLMHPQSMKSDPEKVPVYNA--AAYAR--- 147
Db 654 VTVELRDNDKPKVEKQQLN--NAVSIQNVKPGVT-TDWKETADGVYKATYATYTKSG 710

Qy 148 -----KGFVGQELTQVEMLGTMR-----GVRIALTIN 175
Db 711 LTAKLLMNWNEDLHTAGFIIDANPOSAKIATLSASNGVLNANENANTVSVNVADEGSN 770

Qy 176 PQQYDVV-----ANQLKVRNIEIVSFGQADEVATQRLYDASFSPFYATYKQL-- 225
Db 771 PINDHTVTFAVLSSGSAISFNNQNTAKTDVNLGLATFDLKSQKQEDNTVEVLENGVQTLI 830

Qy 226 --PNRDVYTDHGLYNTPYRMLVAGAKPEALKPMTWAKQGFYL-DVHYT---DEAE 279
Db 831 VSFVGSSTAQVDLQSKNEV---ADGNSVTMTATVRDAKGNLLNDVMVTFNVSAB 886

Qy 280 VGTFTASIKAFIHKYNDGLAASAPVFLALVGDITDVISGEKGGKTKKVTLDLYSADGD 339
Db 887 AKLSQTEVNS-----HDGIATATL-----TSLKNGDY-RVTASVSSSQANQQVN 930

```

| | | | |
|----|------|---|------|
| QY | 340 | YPPEWYTRMSASSPEELTNILDKVLMYEKATMPDKSYLEKVLLIAGADYSWNVSQGQP- | 398 |
| Db | 931 | FIGQSTAALTSLSPSGDITVTNPAQYMATLQDN-----GNPL | 971 |
| QY | 399 | -----TIKCMQYYNQEHGYTDVYNLYKAPTGCY--SHL-----NTGVS | 437 |
| Db | 972 | KXKEITSVPNDNAKSFIS---NGGKGMTDSNGVAJASLTGTLAGTHMIMARLANSVS | 1028 |
| QY | 438 | FANYTAHGSZAWADPLLTTLSOLKALNKDKKYFLAIGNCCICIAQDFVVQPCFGEVITRVK | 497 |
| Db | 1029 | DAQPMTFVADKRVAVVQLTISKAEILGN-----GVDETILTA-----TVK | 1068 |
| QY | 498 | EKGAYAYIGSSPNSYGEDYWSYGANAVFGVQPTFEGSTMGSDATFLDEDSYNTVNSIM | 557 |
| Db | 1069 | DPSNHPVAGITVNFTMPQD---VAANFT-----LENN----- | 1097 |
| QY | 558 | WAGNLAAATHAGNICNIITHIGHYYWEAHYHVLDGDSVMPYRAMPKXTNYITLPASLPQNQAS | 617 |
| Db | 1098 | ---GIAITQANGEAHVTLKG-----RKAGHTTATLGNNTS | 1132 |
| QY | 618 | YS-----IQASAGSVVAISKOGVLXGTGVANASGVATSMTKQITENGNYDVIVTRS | 670 |
| Db | 1133 | DSQPVTFVADKASAOVVLQISKDEIT-NGVDSTAILTATVK-----DOFNEVN | 1180 |
| QY | 671 | YLPVIKOIQVEPSPYOPVSNLTATTQCKVTLK---WEAPSAKKAEGSREVKRIGDGLF | 727 |
| Db | 1181 | NLPV-----TPSSASSGLTLTPGVSNINESGIAQTLAGVAFGEKTV----- | 1222 |
| QY | 728 | VITIEPANDVRANEAKVVLADNWGDNTGQFLLDADHNTEGFSVIPATGPLFTGTASSNL | 787 |
| Db | 1223 | -----TASLANNGASDNKTHFHIGDTAAAKIIELAPVDPDSIIAGT----- | 1262 |
| QY | 788 | YSANFXEYLVPANADPVVTTQNIIVTGQGEVVIPGGVYDYCITNPEPASGKMWLADGGNQ | 847 |
| Db | 1263 | -----PONSSGSVITATVV-----DNNGF | 1281 |
| QY | 848 | PARYDDFTFBAGKKYTFTMR--RAGMDGDTMEVEDDSPASYTYTVYVROGTKIKEGLTAT | 905 |
| Db | 1282 | PVK-----GVTNVFTSNAATAEMINGGOAVINEQGKATVITYVTR--SSIESGARPD | 1331 |
| QY | 906 | TFE---EDGVAAGNHEVCVEKYTAGVSPKVKCDVTVEGSNEFPAPQNLFGS-SVGQKVT | 961 |
| Db | 1332 | TVEASLENGSS-----TLSTSINVNDAS--TAHLTLLQALPDTVSAGETTS | 1376 |
| QY | 962 | LKWDA PNGTPNPENPNPGTLLSESFENGISPASKWTIDADGDGHCKPKGNAPGIAGYN | 1021 |
| Db | 1377 | LYIE-----VKONGVFPQQEVTLTSV-SPSEGVTFESN-NAIYTTN | 1415 |
| QY | 1022 | SNCVKSSEFLOGIGVLTDPNLIITPALDLPNGGKLTFWCQAQDANYASEHYAVIASST | 1081 |
| Db | 1416 | HDCNFYA-SFTATKAGV---YOLTATLE--NGDSMQQTV-TVVPRVNAEITLAASKD | 1466 |
| QY | 1082 | GNDASFNETALLBETITAKGRSPKAIRGBIOGTWRQKTVDLPAGTKYVAFRHFOQSDMF | 1141 |
| Db | 1467 | PVIADN--NDLTTLTATVADTEG-----NAIAANT--EVTFTLEDVK-----ANFT | 1508 |
| QY | 1142 | YIDLDELVEIKANGRADFTETFESSHGEPAPAEWTTIDADGCGWCLCSSGOL----- | 1195 |
| Db | 1509 | LSDGKVIDTAEKGAUKVTLGTXAGAH-----TVTASMTGG-----KSELVVNFIA | 1555 |
| QY | 1196 | DWITAHGSSNVVSSFWMNGMALPNPNLYISKDVTGATKKVYYYAVNDGPPGDHYAVWISK | 1255 |
| Db | 1556 | DLTLAQVNLNUTE-----DNFIANN--VGMTRLQATVTDGNGNPLANEAVTFFIL | 1602 |
| QY | 1256 | TGTVNAGDFTVVFEETPNGINKGARPGLSSTEANGAKPQSVMIBRTVDLPAGTK-----Y | 1309 |
| Db | 1603 | PADVSAFST-----LOQGGSAI--TDINGKAEVTL-----SGTKSGYTPVT | 1641 |
| QY | 1310 | VAFRHNCSDLNYILLDDIQFTWGSSPTPDYTYTVYVROGTKIKEGLTETTEEDGVARG | 1369 |
| Db | 1642 | VSNNYCVSUTKQVTL-----IADAGTAKLASLT-SVYSFVVSFTTEGATMTASVTD--ANG | 1694 |
| QY | 1370 | NHEYCYVEVKYTAGVSPKKCVDVTVNSTQFNPFQNLTAEOAPNSMDAILKNWAP--ASKEA | 1427 |

| | | | | |
|----|------|--|---|------|
| Db | 1695 | NPVEGIKYNF | -----RGTSVILSTSVETDDRGFAEILVTSTEVGLKTVSASLADKPT | 1747 |
| Qy | 1428 | EVLNEDFENGIPASWKTIADGDGNWTTT | --PPPGSSFAGHNSAICVSSASHIN--FE | 1483 |
| Db | 1748 | EVIS | -----RLNADSVNSATITSLLEIPEGVQVMVAQDVAV | 1792 |
| Qy | 1484 | GP | -----QNPDNYLV-----TPE-----LSLPGGGTLTFW | 1508 |
| Db | 1793 | NPVAHQPVTFSAEPSSQMIISQNTVSTNTQGVAEVTTTPERNGSYMKASLPNGASLEKQ | 1852 | |
| Qy | 1509 | VCAQDANY | -----ASEHYAVYASSTGNDASNFANALLEVLTAKTVTVTAPBAIRGTRAQGT | 1564 |
| Db | 1853 | LEAIDEKLLTASSPLIGTVAP | -----TGATLTATLTASNGTPVEGQ | 1894 |
| Qy | 1565 | WYQKTVQLPAGTKY | -----VAFRHFGCTDFFWINLDDVWITSGNAPSVTYTI--YRNNTQIA | 1619 |
| Db | 1895 | VINFSV | --TPEGATLSGGKVRTNSGQA-----PVLTSNKVGTVTVTASPHNGVTIQ | 1945 |
| Qy | 1620 | SGVETTYR | -----DP-----DLAT-----GFYTVGVKVVYPN | 1647 |
| Db | 1946 | ---TQTVKVTGNSSTAHVASFADPSTIAATNTDLSLTKATVEDSGNLEGLTVYF | ---2000 | |
| Qy | 1648 | GESAIETATINITSADVTAQPKYTLFVVGKTIITVTCQGEAMITDMNGRELAAGRNTVYV | 1707 | |
| Db | 2001 | ---ALKGSAATLSLTAVTPDQNGIATTSVKGAMTGSVTVSAV | ---TTAGMGMQIVDI | 2050 |
| Qy | 1708 | TAQGG | 1712 | |
| Db | 2051 | TLVAG | 2055 | |

RESULT 15

| | | | | | |
|-----|---|-----------|------|------|-----|
| OMP | RICCN | STANDARD; | PRT; | 1655 | AA. |
| ID | OMP_RICCN | STANDARD; | PRT; | 1655 | AA. |
| AC | Q9KKA3; Q9KX98; Q9XC45; | | | | |
| DT | 16-OCT-2001 (Rel. 40, Created) | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last sequence update) | | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | | |
| DE | Outer membrane protein B precursor (168 kDa surface-layer protein) (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB) (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide]. | | | | |
| GN | OMP BK OR K1085. | | | | |
| OS | Rickettsia conorii. | | | | |
| OC | Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; | | | | |
| OC | Rickettsiaceae; Rickettsiae; Rickettsia. | | | | |
| OX | NCBI_TaxID=781; | | | | |
| [1] | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=Malish 7; | | | | |
| RC | MEDLINE=1442074; PubMed=11557893; | | | | |
| RA | Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V., | | | | |
| RA | Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., | | | | |
| RA | Raoult D.; | | | | |
| RT | "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."; | | | | |
| RL | Science 293:2093-2098(2001). | | | | |
| [2] | SEQUENCE OF 33-1649 FROM N.A. | | | | |
| RP | STRAIN=Indian tick typhus, and Malish 7; | | | | |
| RC | MEDLINE=20393643; PubMed=10939649; | | | | |
| RA | Roux V., Raoult D.; | | | | |
| RT | "Phylogenetic analysis of members of the genus Rickettsia using the | | | | |
| RT | gene coding the outer-membrane protein rOmpB (ompB)." ; | | | | |
| RL | Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000). | | | | |
| [3] | SEQUENCE OF 353-1655 FROM N.A. | | | | |
| RP | STRAIN=Malish 7; | | | | |
| RC | STRAIN=Malish 7; | | | | |
| RA | Stenos J., Walker D.; | | | | |
| RT | "the rickettsial outer membrane protein A and B genes of Rickettsia | | | | |
| RT | australis, the most divergent rickettsia of the spotted fever group." ; | | | | |
| RL | Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | !- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR | | | | |

STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).

!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).

!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry (By similarity).

!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE008659; AAL03623.1; -
 EMBL; AF123721; AAF34124.1; -
 EMBL; AF123726; AAF34129.1; -
 EMBL; AF149110; AAD39533.1; -
 PIR; E97835; E97835.
 InterPro; IPR006315; Autotransport.
 InterPro; IPR005546; Autotransporter.
 Pfam; PF03797; Autotransporter; 1.
 TIGRFAMs; TIGR01414; autotrans_bar1; 2.
 Antigen; S-layer; Cell wall; Complete proteome.
 CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1335 1655 32 kDa BETA PEPTIDE.
 FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 78 78 K -> S (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 251 251 V -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
 FT CONFLICT 353 354 KD -> GH (IN REF. 3).
 FT CONFLICT 776 776 F -> S (IN REF. 3).
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).
 SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match 2.3%; Score 209.5; DB 1; Length 1655;
 Best Local Similarity 19.2%; Pred. No. 0.0019;
 Matches 277; Conservative 153; Mismatches 501; Indels 511; Gaps 67;

499 KGAYAYIGSSPNSYWGEDIYHVSUGANAVGVQPTFEGTSMG-----SYDA--T 544
 383 KTAASKVAITONSFGTFDGNLAAQIIIVPNTLNGNFTGDASNPGNTAGVITFDANGT 442
 545 FLEDS-----YNTVNSIMWAG-----NLAATHA-----GNIGNITHIGAHYYEAYHVL 588
 443 LASASADANVAVTNNITAIERASGAGVQLSGTHAEELRLNAGSV-----FKL 490
 589 GDGWPMPRAMPKNTYITPLASLPONQASYSIQASAGYVAISKDGVLYGTGVANASGVA 648
 491 ADGTVINGKV-----NOTALVGGALAAAGTITLDGSATITG-DIGNAGGAA 534
 649 TV-----SMTKOIT-----ENGYDVVITRSNVLPIKIQIQVGP 683
 535 ALQGITLANTAKITLUGGANIIGANGGTINFQANGGTIKLTSTQNN---IVVDFDLAIA 591
 684 SPYQFVSNLTATQCKVTLKWEAPSAKKAGSREVKRGDGLFVTIETEPANDVRANEAKV 743
 592 TDQGVVDASSLTWAQTLTINGKIGTV--GANNKTLGQFNIGSSKTVLSGDVAINEI-- 647
 744 VLADNWGDNTGQFLLDADHNT-----FGSVIPATGPLFTGT---A 783
 648 -----VIGNNGAVQFA---HNTYLIPTRTTNAAGQGGKIIIFNPVVNNNTTLATGTLGSA 697
 784 SSNLYSANFEYLVFANADPVVITQNIIVTGGGEVVIPIGVVDYCIITPEP-----AS 835

698 TNPLAEINFGSKGAANDVTLN-----VGKV-----NLVATNITTTDANVGSIINAG 746
 836 GKMWIAGD-GGNQPARYDDFTFEAGKYFTFMRBAGMGDGDMEVEDDSPAITYTT----- 890
 747 GTNIVSGTVGGQQGNKENTVALDNGTTVKF-LGNATFNGNTTIAANSTLQIGGNYTADFV 805
 891 VYRDGTKIKKGLTATTFFEEGVAAGNHEYCVVKYTAGVSP-KVKCKDVTVEGS-----NE 944
 806 ASADGTGLIVE-----FVNTGP-----ITVTLNKAAPFVNALKQITVSGPGNVVINE 851
 945 FAPVQNLTG-----SSVGOKVTLKWDAPNGTP-NPNPNPNP-----NPGTTLSES 988
 852 IGNAAGNYHGAVTDTIAFENSSLGAUVFL---PRGIFPDAGNRILPLTIKSTGVGNKATG 907
 989 PE--NGIPASWKTIDADGDGHGKPGNAPGIAGNNSGCVVYSFGLGGIGLVLPDNYLI 1046
 908 FDESVIVLGVDSVIADQVIG-DQNNIVGLGLSGDNDIIVNATTLVAGIGTINNQGTV 966
 1047 TPALDLNPGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF-----T 1089
 967 TLSGGIEN-----TPGTVYGLGTGIGASKFKQVTTTDDYNNIGNIAT 1009
 1090 NALLEE--TITAKGR---SPKAIRGRIQOTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 1143
 1010 NATINDGVTVTGGIAGIGFDGKITLGSVNGNVRFD----- 1048
 1144 DLDEVEIKANGKRAADFETTESSTHGEAPAEWTTIDADGGQGMCLSSGQDLWLTAGHG 1203
 1049 -----GILSHSTSMIGTTKANNG 1066
 1204 SNVVSFSFSGMALNPDNYLISKDVATKVKYVYVNDGFPDGHYAVMLSKTGTNAG-- 1261
 1067 TVTVLGNAPVGNIGSDSTPVASVFTGSDG-----GAGLOGNIYQVDFDFTYINLGIS 1119
 1262 DFTVVFETENGIN-KGARFPLSTANGAKPOSVW-----IBRTVDLPAGTKYVAFRHY 1315
 1120 NSNVILGGGTAINGKINLNTLTTFASGT---STWGNNTSIETTLTANG----- 1167
 1316 NCSDLNVLDDDTQFTWGGSPPT-----DYTYTVYRDGTKIKEGLTETPTE 1362
 1168 NIG--NIVILEGAGVNAATTGTTTIKVQDNANANFSGTQTYTLIQGARENGTLGGPNF- 1224
 1363 EDGVATGNHEYCVVEKYTAGVSPKKCVDVTVNSQFNPQNLTAEOAPNSMDALLKWNAP 1422
 1225 ---VVTGSNRF---VNY--GLIRAAQDYVI--TRTNNAENVVTN----- 1259
 1423 ASKRAEVLNEDFPENGIPASWKTIDADGGNNWTTTPPGSSSPAGHNSAICVSSASHNF 1482
 1260 -----DIANSF-SGAP-----GVGQN-VTTFVNATTAAYNNLLILAKNSANSANF 1303
 1483 EGPQNDP-----NYLAVTP---ELSLPGGGTTLTFWYCAQ 1512
 1304 VGLAVTDTSAITNAQLDVAKDIOAQLGNLGAIRYLCTPETAEAGPEAGAIIPAAVAAG 1363
 1513 DANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRAGTQWYQKTVQL 1572
 1364 DEADVNVAYGIWAKPFYVTDHQ-----SKKGLAGYKAKTT----- 1399
 1573 PAGTKYVAFRHFCTGTFFWINLDDVITSGNAPSYYTYIYRNNTQI--ASGVTEYTRDP 1630
 1400 -----GVV-----IGLD-----TLANDNLMIAGAITKTIDIKHQ 1429
 1631 DLATG-----FVTYG---VKVYPNGESAIEATATLINITSLADVTAQKPYTLTVVG 1677
 1430 DYKKGDKTVNGSFSLYGAQQLVQKFAQG-SAI--FSLN----- 1467
 1678 KTIIVTQCGBAMTYDMNG---RELAAGR-----TVVYTAQGHYAVVWVVDGKSY 1725
 1468 ---QVKNSQRVYFDANGNMSKQIAAGHYDNMTFGNLTGVYDYNAMQGVLVTPMAGLSY 1524
 1726 VE 1727

Db 1525 LK 1526

Search completed: May 18, 2004, 11:33:32
Job time : 35.7568 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:28:29 ; Search time 69.1557 Seconds
(without alignments)
7902.136 Million cell updates/sec

Title: US-08-353-485-10

Perfect score: 9179

Sequence: 1 MKLLLLIAASLLGVLYAQ.....HYAVNVVDGKSYVEKLVK 1732

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------|--------------------|
| 1 | 9179 | 100.0 | 1732 | Q51817 | Q51817 porphyromon |
| 2 | 9167 | 99.9 | 1732 | O07442 | O07442 porphyromon |
| 3 | 9158 | 99.8 | 1732 | O52050 | O52050 porphyromon |
| 4 | 8029.5 | 87.5 | 1723 | P72194 | P72194 porphyromon |
| 5 | 7774.5 | 84.7 | 1723 | P72197 | P72197 porphyromon |
| 6 | 6486 | 70.7 | 1223 | Q9ZNB5 | Q9ZNB5 porphyromon |
| 7 | 6156.5 | 67.1 | 1358 | P96967 | P96967 porphyromon |
| 8 | 4272 | 46.5 | 1706 | Q51839 | Q51839 porphyromon |
| 9 | 4270 | 46.5 | 1704 | Q51816 | Q51816 porphyromon |
| 10 | 4256 | 46.4 | 1706 | Q51838 | Q51838 porphyromon |
| 11 | 4249.5 | 46.3 | 1097 | P72196 | P72196 porphyromon |
| 12 | 4245 | 46.2 | 1687 | Q9R9B7 | Q9R9B7 porphyromon |
| 13 | 1854 | 20.2 | 364 | P96966 | P96966 porphyromon |
| 14 | 980.5 | 10.7 | 925 | Q9FAJ0 | Q9FAJ0 porphyromon |
| 15 | 722 | 7.9 | 312 | Q9KLB3 | Q9KLB3 porphyromon |
| 16 | 547 | 6.0 | 736 | Q51844 | Q51844 porphyromon |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 17 | 541 | 5.9 | 736 | 2 | O33441 | O33441 porphyromon |
| 18 | 354.5 | 3.9 | 293 | 2 | Q9XB09 | Q9XB09 porphyromon |
| 19 | 341.5 | 3.7 | 422 | 2 | Q51818 | Q51818 porphyromon |
| 20 | 322 | 3.5 | 5291 | 16 | Q8X2T1 | Q8X2T1 escherichia |
| 21 | 315 | 3.4 | 5188 | 16 | Q8X4H5 | Q8X4H5 escherichia |
| 22 | 313 | 3.4 | 1993 | 16 | Q92FG5 | Q92FG5 listeria in |
| 23 | 309 | 3.4 | 1744 | 16 | Q82YW8 | Q82YW8 enterococcu |
| 24 | 306.5 | 3.3 | 2219 | 16 | Q88W19 | Q88W19 lactobacill |
| 25 | 303 | 3.3 | 4210 | 16 | Q89J13 | Q89J13 bradyrhizob |
| 26 | 300.5 | 3.3 | 2468 | 16 | Q9I2M3 | Q9I2M3 pseudomonas |
| 27 | 300 | 3.3 | 8173 | 16 | Q7UDU8 | Q7UDU8 rhodopirell |
| 28 | 295.5 | 3.2 | 1817 | 17 | Q8T159 | Q8T159 methanosarc |
| 29 | 295.5 | 3.2 | 3346 | 16 | Q7WN54 | Q7WN54 bordetella |
| 30 | 295.5 | 3.2 | 4199 | 16 | P74440 | P74440 synechocyst |
| 31 | 295 | 3.2 | 1975 | 2 | Q8GJ03 | Q8GJ03 enterococcu |
| 32 | 294 | 3.2 | 2522 | 16 | Q8EKA6 | Q8EKA6 shewanella |
| 33 | 291 | 3.2 | 2523 | 17 | Q8TJ88 | Q8TJ88 methanosarc |
| 34 | 289 | 3.1 | 3988 | 17 | Q8TP21 | Q8TP21 methanosarc |
| 35 | 285 | 3.1 | 1861 | 16 | Q89L34 | Q89L34 bradyrhizob |
| 36 | 284.5 | 3.1 | 1752 | 2 | Q9AE52 | Q9AE52 ruminococcu |
| 37 | 283 | 3.1 | 2566 | 17 | Q8TSE7 | Q8TSE7 methanosarc |
| 38 | 280.5 | 3.1 | 3029 | 16 | Q55582 | Q55582 synechocyst |
| 39 | 280 | 3.1 | 2016 | 17 | Q8TJ89 | Q8TJ89 methanosarc |
| 40 | 280 | 3.1 | 2230 | 16 | Q7U7J7 | Q7U7J7 synechococc |
| 41 | 279.5 | 3.0 | 8682 | 16 | Q88RG2 | Q88RG2 pseudomonas |
| 42 | 278.5 | 3.0 | 1541 | 16 | Q7UQU9 | Q7UQU9 rhodopirell |
| 43 | 278.5 | 3.0 | 6310 | 16 | Q88PP2 | Q88PP2 pseudomonas |
| 44 | 276.5 | 3.0 | 2044 | 16 | Q8Y8Q4 | Q8Y8Q4 listeria mo |
| 45 | 274.5 | 3.0 | 1673 | 1 | Q977V5 | Q977V5 methanosarc |

ALIGNMENTS

RESULT 1

Q51817 ID Q51817 PRELIMINARY; PRT; 1732 AA.
AC Q51817;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Porphyropain.
GN PRTP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_taxid=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W12;
RX MEDLINE=96213011; PubMed=8631659;
RA Bartoccy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA Progulskis-Fox A., Lantz M.S.;
RT "Analysis of the prtp gene encoding porphyropain, a cysteine proteinase
of Porphyromonas gingivalis.";
RL J. Bacteriol. 178:2734-2741 (1996).
DR EMBL; U42210; AAB06565.1; --
DR PIR; T30836; T30836.
DR MEROPS; C25.002; --
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0008234; P:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; P:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25_I_1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCA4 CRC64;

| | | | | | | |
|-----------------------|-----|---|---------------------------------|-----------|--------------|--|
| Query Match | | 100.0%; | Score 9179; | DB 2; | Length 1732; | |
| Best Local Similarity | | 100.0%; | Pred. No. 0; | | | |
| Matches 1732; | | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; | |
| QY | 1 | MRKLLIIAASLGVGLVYASAKIKLDA | PTTRTTCTNNSPKQFDASFSNEVELTKVETK | 60 | | |
| DB | 1 | MRKLLIIAASLGVGLVYASAKIKLDA | PTTRTTCTNNSPKQFDASFSNEVELTKVETK | 60 | | |
| QY | 61 | GGTFASVISIPGAPTEGVGSPVPAVRKLTAVPVGATPVVRVKSFTQVYSLNQYSEKL | 120 | | | |
| DB | 61 | GGTFASVISIPGAPTEGVGSPVPAVRKLTAVPVGATPVVRVKSFTQVYSLNQYSEKL | 120 | | | |
| QY | 121 | MPHQPSMSKDDPEKVPFVYVNAAYARKFVGQELTQVEMGLTMRGVRIIAALINPVQYD | 180 | | | |
| DB | 121 | MPHQPSMSKDDPEKVPFVYVNAAYARKFVGQELTQVEMGLTMRGVRIIAALINPVQYD | 180 | | | |
| QY | 181 | VVANQLKVRNNIEIVSFGQADEVATQRLYDASFSFYFETAYKQLENRDVYTDHGDLYNT | 240 | | | |
| DB | 181 | VVANQLKVRNNIEIVSFGQADEVATQRLYDASFSFYFETAYKQLENRDVYTDHGDLYNT | 240 | | | |
| QY | 241 | PVRMLVVAGAKFKEALKPMLTWAQKGFYLDVHYTDEAEVGTNNASIKAFIHKKYNDGLA | 300 | | | |
| DB | 241 | PVRMLVVAGAKFKEALKPMLTWAQKGFYLDVHYTDEAEVGTNNASIKAFIHKKYNDGLA | 300 | | | |
| QY | 301 | ASAPVFLALVGDPTVISGEGKKTKKVTDLYYSANDGYFFPEMYTFRMSASSPEELTNI | 360 | | | |
| DB | 301 | ASAPVFLALVGDPTVISGEGKKTKKVTDLYYSANDGYFFPEMYTFRMSASSPEELTNI | 360 | | | |
| QY | 361 | IDKVLMEKATMPDKSYLEKVLIIAGADYSNWSQVGQPTTIKYGQVYVNOEHGYTDVINY | 420 | | | |
| DB | 361 | IDKVLMEKATMPDKSYLEKVLIIAGADYSNWSQVGQPTTIKYGQVYVNOEHGYTDVINY | 420 | | | |
| QY | 421 | LKAPYTCYSHLNTGVSFANYTAHGETAWADPLLTSQKALTNKDKYFLAIGNCCITA | 480 | | | |
| DB | 421 | LKAPYTCYSHLNTGVSFANYTAHGETAWADPLLTSQKALTNKDKYFLAIGNCCITA | 480 | | | |
| QY | 481 | QFYDVPQCFGEVITRVEKEGAYAYIGSSPNSYWGEDIYWSVGANAVFGVQPTFEGTSMGS | 540 | | | |
| DB | 481 | QFYDVPQCFGEVITRVEKEGAYAYIGSSPNSYWGEDIYWSVGANAVFGVQPTFEGTSMGS | 540 | | | |
| QY | 541 | YDATFELDSYNTVNSIMWAGNLAATHAGNIGNITHGAHYWYAYHVLGDSYMPYRAMP | 600 | | | |
| DB | 541 | YDATFELDSYNTVNSIMWAGNLAATHAGNIGNITHGAHYWYAYHVLGDSYMPYRAMP | 600 | | | |
| QY | 601 | KINTYTLPASLPQNASYSIQASAGSYVAISKDGLVYGTGVANASGVATVSMTKQITENG | 660 | | | |
| DB | 601 | KINTYTLPASLPQNASYSIQASAGSYVAISKDGLVYGTGVANASGVATVSMTKQITENG | 660 | | | |
| QY | 661 | NYDWTIRSNYLPVVKIQIOVGEPSPYQPVSNLTATTOGQKVTLKWEAPSAKKAEGGREVK | 720 | | | |
| DB | 661 | NYDWTIRSNYLPVVKIQIOVGEPSPYQPVSNLTATTOGQKVTLKWEAPSAKKAEGGREVK | 720 | | | |
| QY | 721 | RIGDGLFVTTEPANDVRANEAKVLAADNVWGNNTGYQFLLDADHNTFGSVIPATGPLEFT | 780 | | | |
| DB | 721 | RIGDGLFVTTEPANDVRANEAKVLAADNVWGNNTGYQFLLDADHNTFGSVIPATGPLEFT | 780 | | | |
| QY | 781 | GTASSNLYSANFEYLVPANADPVVTTQNIIVTGGGEVVIIPGVVYDICTINPEPASGQWMI | 840 | | | |
| DB | 781 | GTASSNLYSANFEYLVPANADPVVTTQNIIVTGGGEVVIIPGVVYDICTINPEPASGQWMI | 840 | | | |
| QY | 841 | ADGGNQPARYDDFTFAGKYYFTMRAGMGDGTDMVEVDDSPASVYTYVYRDGTIKE | 900 | | | |
| DB | 841 | ADGGNQPARYDDFTFAGKYYFTMRAGMGDGTDMVEVDDSPASVYTYVYRDGTIKE | 900 | | | |
| QY | 901 | GLTATTEEDGVAGNHEYCEVVKYTAGVSPKCVKDTTVVSGSNEFAFPVQNLTGSSVGOKY | 960 | | | |
| DB | 901 | GLTATTEEDGVAGNHEYCEVVKYTAGVSPKCVKDTTVVSGSNEFAFPVQNLTGSSVGOKY | 960 | | | |
| QY | 961 | TLKWDAPNGTNPFPNPNPFGTTLSSEFENGIPASWKTIDADGDGHGKPGNAPGLAGY | 1020 | | | |
| DB | 961 | TLKWDAPNGTNPFPNPNPFGTTLSSEFENGIPASWKTIDADGDGHGKPGNAPGLAGY | 1020 | | | |

| | | | |
|----|------|---|------|
| QY | 1021 | NSNGCVYSSEFGLGGIGVLTPDNYLITPDLDPNGGKLTFWVCAQDANYASHEHYAYASS | 1080 |
| DB | 1021 | NSNGCVYSSEFGLGGIGVLTPDNYLITPDLDPNGGKLTFWVCAQDANYASHEHYAYASS | 1080 |
| QY | 1081 | TGNDASNFNTALLEETITAKGVRSKPAIRGRIOGTWRQKTVDLPAGTKVYVAFRHPOSTDM | 1140 |
| DB | 1081 | TGNDASNFNTALLEETITAKGVRSKPAIRGRIOGTWRQKTVDLPAGTKVYVAFRHPOSTDM | 1140 |
| QY | 1141 | FYIDLDEVEIKANGKRADETFETFESSTHGEAPAEWTTIIDADGGQGWLCGLSSQLDLWLT | 1200 |
| DB | 1141 | FYIDLDEVEIKANGKRADETFETFESSTHGEAPAEWTTIIDADGGQGWLCGLSSQLDLWLT | 1200 |
| QY | 1201 | HGGSNVVSSFSWMGNALNPDNYLISKDVTGATKVYKYIYVNDGFPDGDHVAVMISKTGTNA | 1260 |
| DB | 1201 | HGGSNVVSSFSWMGNALNPDNYLISKDVTGATKVYKYIYVNDGFPDGDHVAVMISKTGTNA | 1260 |
| QY | 1261 | GDFTVVFEETPNGINKGGARFGLSTEANGAKPOSWIERTVDLPAGTKVYVAFRHYNCSDL | 1320 |
| DB | 1261 | GDFTVVFEETPNGINKGGARFGLSTEANGAKPOSWIERTVDLPAGTKVYVAFRHYNCSDL | 1320 |
| QY | 1321 | NYILLDDIOFTWGGSPPTDYTVTVYRDGTIKKEGLTETTFEEDGVATGNHEYCEVVKYT | 1380 |
| DB | 1321 | NYILLDDIOFTWGGSPPTDYTVTVYRDGTIKKEGLTETTFEEDGVATGNHEYCEVVKYT | 1380 |
| QY | 1381 | AGVSPKCCVDVTNVSQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA | 1440 |
| DB | 1381 | AGVSPKCCVDVTNVSQFNPVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA | 1440 |
| QY | 1441 | SWKTTDADGGNNWTTTPPGSSFAGHNSATCVSSASHINPEGPONPNYLVTPELSIP | 1500 |
| DB | 1441 | SWKTTDADGGNNWTTTPPGSSFAGHNSATCVSSASHINPEGPONPNYLVTPELSIP | 1500 |
| QY | 1501 | GGTTLFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEALRGTR | 1560 |
| DB | 1501 | GGTTLFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEALRGTR | 1560 |
| QY | 1561 | AQGTWYQKTVQVLPAGTKYVAFRHFGCTDFFWINLDDVITSGNAPSYYTIVRNNTQIAS | 1620 |
| DB | 1561 | AQGTWYQKTVQVLPAGTKYVAFRHFGCTDFFWINLDDVITSGNAPSYYTIVRNNTQIAS | 1620 |
| QY | 1621 | GVTEITVYRDPDLATGFTYGVKVYVNGESAELETATLNTISLADVTAQKPYTLTVVGKTI | 1680 |
| DB | 1621 | GVTEITVYRDPDLATGFTYGVKVYVNGESAELETATLNTISLADVTAQKPYTLTVVGKTI | 1680 |
| QY | 1681 | TVTCQGEAMIDYMNRRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK | 1732 |
| DB | 1681 | TVTCQGEAMIDYMNRRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK | 1732 |

RESULT 2

007442 PRELIMINARY; PRT; 1732 AA.
AC 007442;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Lysine-specific cysteine proteinase.
GN PRK.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=99235907; PubMed=10219167;
RA Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT "Characterization of a Porphyromonas gingivalis gene prk that encodes
a lysine-specific cysteine proteinase and three sequence-related
adhesins."
RL Oral Microbiol. Immunol. 14:92-97 (1999).
DR EMBL: U75366; AAB60809.1; --
DR MEROPS: C25.002; --

| | | | |
|------|----|---|------|
| 781 | Db | GTASSNLSANFELYLIPANADPVVVTTQNIITVTCQGEVVIPEGVYDYCIITNPPEASGQWMI | 840 |
| 841 | Qy | AGDGNOPARYDDFTPEAGKKYFTWRRAGWGDTGMEVEDDPSASYTYTVYRDGTKE | 900 |
| 841 | Db | AGDGNOPARYDDFTPEAGKKYFTWRRAGWGDTGMEVEDDPSASYTYTVYRDGTKE | 900 |
| 901 | Qy | GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVKDVTVEGSEFAFPVQNLTGSSVGQKV | 960 |
| 901 | Db | GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVKDVTVEGSEFAFPVQNLTGSSVGQKV | 960 |
| 961 | Qy | TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTTDADGDGHGKPGNAPGIAGY | 1020 |
| 961 | Db | TLKWDAPNGTPNPNPNPNPGTTLSESFENGIPASWKTTDADGDGHGKPGNAPGIAGY | 1020 |
| 1021 | Qy | NSNGCVYSEFGLGGIGLVLTDPNKLTPALDLPNGGLTFWVCAQDANYASEHYAVYASS | 1080 |
| 1021 | Db | NSNGCVYSEFGLGGIGLVLTDPNKLTPALDLPNGGLTFWVCAQDANYASEHYAVYASS | 1080 |
| 1081 | Qy | TGNDASFTNALLEETITAKGVSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM | 1140 |
| 1081 | Db | TGNDASFTNALLEETITAKGVSPKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM | 1140 |
| 1141 | Qy | FYIDLDEVEIKANGKRAFFTETPESSTHGAPAEWTTIDADGGQGWLCUSSQLDWLTA | 1200 |
| 1141 | Db | FYIDLDEVEIKANGKRAFFTETPESSTHGAPAEWTTIDADGGQGWLCUSSQLDWLTA | 1200 |
| 1201 | Qy | HGGSNVVSSFWNGMALPNPNYLISKDVTGATKVXYIYVNDGFGPDHYAVMLSKGTGNA | 1260 |
| 1201 | Db | HGGSNVVSSFWNGMALPNPNYLISKDVTGATKVXYIYVNDGFGPDHYAVMLSKGTGNA | 1260 |
| 1261 | Qy | GDTVVPFEETPNGIKGGARFGLSTEANGAKPOSWIERTVDLPAGTKYVAFRHYNCSDL | 1320 |
| 1261 | Db | GDTVVPFEETPNGIKGGARFGLSTEANGAKPOSWIERTVDLPAGTKYVAFRHYNCSDL | 1320 |
| 1321 | Qy | NYILLDDIQHTGGSPPTDITYTVYTRDGTKIKEGLTETTFEEDGVTATGNHEYCVEVKYT | 1380 |
| 1321 | Db | NYILLDDIQHTGGSPPTDITYTVYTRDGTKIKEGLTETTFEEDGVTATGNHEYCVEVKYT | 1380 |
| 1381 | Qy | AGVSPKKCVDTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKAEVLNEDFENGIPA | 1440 |
| 1381 | Db | AGVSPKKCVDTVNSTQFNPVQNLTAEQAPNSMDAILKWNAPASKAEVLNEDFENGIPA | 1440 |
| 1441 | Qy | SWKTTIDADGDNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQPNLYLVTPELSLP | 1500 |
| 1441 | Db | SWKTTIDADGDNWTTTPPPGGSSFAGHNSAICVSSASHINFEGPQPNLYLVTPELSLP | 1500 |
| 1501 | Qy | GGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANALLEEVLTAKTVVTPEAIRGTR | 1560 |
| 1501 | Db | GGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANALLEEVLTAKTVVTPEAIRGTR | 1560 |
| 1561 | Qy | AQGTWTKQTVLPAGTKYVAFRHFQCTDFPINLDDVLTSGNAPSYTYTIRYNNNTQIAS | 1620 |
| 1561 | Db | AQGTWTKQTVLPAGTKYVAFRHFQCTDFPINLDDVLTSGNAPSYTYTIRYNNNTQIAS | 1620 |
| 1621 | Qy | GVTETTYRDPDLATGYTYGVKVVYVNGESAIEATLNTITSLADVTAQKPYTLTVVGKTI | 1680 |
| 1621 | Db | GVTETTYRDPDLATGYTYGVKVVYVNGESAIEATLNTITSLADVTAQKPYTLTVVGKTI | 1680 |
| 1681 | Qy | TVTTCQGEAMIDMNGRRLAAGRNVTVVYTAQGGHYAVMVVDGKSYVEKLAVK | 1732 |
| 1681 | Db | TVTTCQGEAMIDMNGRRLAAGRNVTVVYTAQGGHYAVMVVDGKSYVEKLAVK | 1732 |

06. Created)

, Last sequence update)

, Last annotation update

Bacteroides gingivalis)
rotease.

acteroides (class); Bacteroidales;

OC Porphyromonadaceae; Porphyromonas.
 NCBI_TaxID=837;
 [i]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=98298016; PubMed=9632563;
 RT "Is195, an insertion sequence-like element associated with protease
 genes in Porphyromonas gingivalis";
 RL Infect. Immun. 66:3035-3042(1998).
 DR EMBL; AF017059; AAC26523.1; -.
 DR MEROPS; C25.002; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1732 AA; 187931 MW; B2337463D5CB5EA5 CRC64;

 Query Match 99.8%; Score 9158; DB 2; Length 1732;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1729; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

 QY 1 MRKLLIIAASLLGVLGYAOSAKIKLDAPTRTCTNNKSKQDPASFSNEVELTKVETK 60
 DB 1 MRKLLIIAASLLGVLGYAOSAKIKLDAPTRTCTNNKSKQDPASFSNEVELTKVETK 60

 QY 61 GGTFAVSIPCAPPTGVSPEVPAVKLLIAPVVGATPVVRVKSFTBQVYSLNQYSGSEK 120
 DB 61 GGTFAVSIPCAPPTGVSPEVPAVKLLIAPVVGATPVVRVKSFTBQVYSLNQYSGSEK 120

 QY 121 MPHQPSKSDDBEKVPFVNAAYARKFGVQBLTQVEMLGTWRGVRIIAALTINPVQYD 180
 DB 121 MPHQPSKSDDBEKVPFVNAAYARKFGVQBLTQVEMLGTWRGVRIIAALTINPVQYD 180

 QY 181 VVANQLKVRNNIELEVSFQGADEVATQRLYDASFSFYFETAYKQLENRVVYTHGDLVNT 240
 DB 181 VVANQLKVRNNIELEVSFQGADEVATQRLYDASFSFYFETAYKQLENRVVYTHGDLVNT 240

 QY 241 PVRLVAVAGAKFKALKPWLTKAQKGYLDVHYTDEAEVGTNNASIKAFIHKYNDGLA 300
 DB 241 PVRLVAVAGAKFKALKPWLTKAQKGYLDVHYTDEAEVGTNNASIKAFIHKYNDGLA 300

 QY 301 ASAPVFLALVGDVTLVSGEGKTKTKVTDLYSAVDGDFPEMYTPRMSASSPEELTNI 360
 DB 301 ASAPVFLALVGDVTLVSGEGKTKTKVTDLYSAVDGDFPEMYTPRMSASSPEELTNI 360

 QY 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSNWSQVGPQTIKYGQYYNNQEHGTVYVNY 420
 DB 361 IDKVLMEKATMPDKSYLEKVLIIAGADYSNWSQVGPQTIKYGQYYNNQEHGTVYVNY 420

 QY 421 LKAPYTCGYSHLNTGVSFANTHAGSETAWADPLLTTSQKALTNKOKYFLAIGNCCITA 480
 DB 421 LKAPYTCGYSHLNTGVSFANTHAGSETAWADPLLTTSQKALTNKOKYFLAIGNCCITA 480

 QY 481 QFDVYVQPCFGEVITRVKEKGAAYVIGSSPNSYWGDEYVWSVGANAVGVQTPFGTSMGS 540
 DB 481 QFDVYVQPCFGEVITRVKEKGAAYVIGSSPNSYWGDEYVWSVGANAVGVQTPFGTSMGS 540

 QY 541 YDAITFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGVSVMPEYRAMP 600
 DB 541 YDAITFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGVSVMPEYRAMP 600

 QY 601 KNTNTYTLPASLPQNOASYSIQASAGSYVAISKDGLYGTGVANASGVATVSMTKQITENG 660

DB 601 KNTNTYTLPASLPQNOASYSIQASAGSYVAISKDGLYGTGVANASGVATVSMTKQITENG 660
 QY 661 NYDVWITRSNYLPIVKIOIQVGEPSPIQVSNLTATTQGGKVLTKWEPAPAKAAGSREVK 720
 DB 661 NYDVWITRSNYLPIVKIOIQVGEPSPIQVSNLTATTQGGKVLTKWEPAPAKAAGSREVK 720
 QY 721 RIDGGLFVITIEPANDVRANEAKVLAADNVWGNTGYQFLLDADHNTFGSVIPATGPLEFT 780
 DB 721 RIDGGLFVITIEPANDVRANEAKVLAADNVWGNTGYQFLLDADHNTFGSVIPATGPLEFT 780
 QY 781 GTASSNLYSANFEYLVPANADPVVTQNIIVTQGGSEWIPGGVYDYCIITNPEPASGKMWI 840
 DB 781 GTASSNLYSANFEYLVPANADPVVTQNIIVTQGGSEWIPGGVYDYCIITNPEPASGKMWI 840
 QY 841 AGDGGNQPARYDDFTFEAGKKYTFTRRAGMGDGTMEVEDDPSASYTTVYVYRDGCKIKE 900
 DB 841 AGDGGNQPARYDDFTFEAGKKYTFTRRAGMGDGTMEVEDDPSASYTTVYVYRDGCKIKE 900
 QY 901 GLTATTFEEDGVAAGNHEVCVEVKYTAGVSPKCKDVTVEGSENEFAPVQNLTGSSVGQKV 960
 DB 901 GLTATTFEEDGVAAGNHEVCVEVKYTAGVSPKCKDVTVEGSENEFAPVQNLTGSSVGQKV 960
 QY 961 TLKWDAPNGTNPNNPNPNTLSESPENGIPASWKTIDADGCHGWKPGNAPGIAGY 1020
 DB 961 TLKWDAPNGTNPNNPNPNTLSESPENGIPASWKTIDADGCHGWKPGNAPGIAGY 1020
 QY 1021 NSNGCVYSSEFGIGGLVLTDPNLYLITPALDLPNGGLTFWVCAQDANYASHEHYAVASS 1080
 DB 1021 NSNGCVYSSEFGIGGLVLTDPNLYLITPALDLPNGGLTFWVCAQDANYASHEHYAVASS 1080
 QY 1081 TGNDAFNFNALLEETITAKGVSPKAIIRGRIOGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
 DB 1081 TGNDAFNFNALLEETITAKGVSPKAIIRGRIOGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
 QY 1141 FYIDLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTIIDADGGQGWLCSSQLDWLTA 1200
 DB 1141 FYIDLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTIIDADGGQGWLCSSQLDWLTA 1200
 QY 1201 HGSNNVSSFSWNGMALNPNDNYLISKDVTGATKVKYVYANDGFGPDGHVAVMLSKTGTA 1260
 DB 1201 HGSNNVSSFSWNGMALNPNDNYLISKDVTGATKVKYVYANDGFGPDGHVAVMLSKTGTA 1260
 QY 1261 GDTVTVFEETPNKINGKGFGLSTANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDL 1320
 DB 1261 GDTVTVFEETPNKINGKGFGLSTANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDL 1320
 QY 1321 NYILLDDIQFTMGSSPTDPTVTVYRDGTHIKEGLTETTTFEEDGVATGNHVEYKVT 1380
 DB 1321 NYILLDDIQFTMGSSPTDPTVTVYRDGTHIKEGLTETTTFEEDGVATGNHVEYKVT 1380
 QY 1381 AGVSPKKCVDTVNSTQFNPVQNLTAEOAPNSMDAILKNNAPAKRAEVLNEDFENGIPA 1440
 DB 1381 AGVSPKKCVDTVNSTQFNPVQNLTAEOAPNSMDAILKNNAPAKRAEVLNEDFENGIPA 1440
 QY 1441 SWKTIDADGDNWTTTTPPGSSFAGHNSAICVSSASHINFEQPNPNLYVTPELSPL 1500
 DB 1441 SWKTIDADGDNWTTTTPPGSSFAGHNSAICVSSASHINFEQPNPNLYVTPELSPL 1500
 QY 1501 GGTLLFWVCAQDANYASHEHYAVASSTGNDASNANLLEEVLTAKTVVTAPEAIRGTR 1560
 DB 1501 GGTLLFWVCAQDANYASHEHYAVASSTGNDASNANLLEEVLTAKTVVTAPEAIRGTR 1560
 QY 1561 AQGTWTKVQLPAGTKYVAFRHFQCTDFWNLDDWITSGNAPSYTYTYIRNNTQIAS 1620
 DB 1561 AQGTWTKVQLPAGTKYVAFRHFQCTDFWNLDDWITSGNAPSYTYTYIRNNTQIAS 1620
 QY 1621 GVTETTYRDPDLATGTYTGKVVYPNGESAIFETATLNTSLADVTQAQPYTLTVVGKTI 1680
 DB 1621 GVTETTYRDPDLATGTYTGKVVYPNGESAIFETATLNTSLADVTQAQPYTLTVVGKTI 1680
 QY 1681 TVTCQGEAMTYDNWGRRLAAGRNTVTVTAQGHYAMVVDGKSYVEKLAVK 1732


```

QY 1261 GDTTVVFEETPENGKGGARFGLSTANGAKPOSVWIERVDPAGTKYVAFRHHYNCSDL 1320
Db 1260 GDTTVVFEETPENGKGGARFGLSTANGAKPOSVWIERVDPAGTKYVAFRHHYNCSDL 1319
QY 1321 NYLLDDIIOFTMGSGPTDPTDYTVYVRDGTGKKEGLTETTFEEDGVATGNHGYCUEVKYT 1380
Db 1320 NYLLDDIIOFTMGSGPTDPTDYTVYVRDGTGKKEGLTETTFEEDGVATGNHGYCUEVKYT 1379
QY 1381 AGVSPKKCVDTVNSTQPNVQNLTAEOAPNSMDAILKNWAPASKRAE 1428
Db 1380 AGVSPKKCVDTVNSTQPNVQNLTAEOAPNSMDAILKNWAPASKRAE 1437
QY 1429 -----VLNEDFENGIPASWKITIDADGNGNWTTPPGSGSFAGHNSA 1471
Db 1438 FVIERPANDVRANEAKVLLAADNVGNTGYQL-LDADHNTFGSVIPATGFLTFASS 1496
QY 1472 ICVSSASHINFEG--PQNP-----NYLVT--PELSLPGGGLTFWVCAQDANYASEHY 1521
Db 1497 NLYSA---NFEVLIPANADPVVTTQNLIVTGQGVVPG---VVDYCIITNEPASGKM 1549
QY 1522 AVYASGTGDASFNANALBEVLTAKTIVVTPAIRGTRAQGTWYQKTVQLPAGTKYV-A 1580
Db 1550 WI-AGDGNQPARYDDTFE-----AGKXYTET 1576
QY 1581 FRHFGCTDFPINLDVWLTSGNAPSVTYTYXNNLTQIASGVTTETTYROPDLATGFYTG 1640
Db 1577 MRRAGMDG-----TMEVEDDPSATYTYVYRDGTGKKEGLTETTYRDMGSAQHEYC 1631
QY 1641 VKVYPNGESALETATINITSADVTAQKPYTLTVVGKTTITVTCOGEAMIDMNGRRLAA 1700
Db 1632 VEKVAAGSPKVCVDYIPDGVADVTAKPYTLTVVGKTTITVTCOGEAMIDMNGRRLAA 1691
QY 1701 GRNTVYVTAQGHYAVWVVDGKSVEKLAVK 1732
Db 1692 GRNTVYVTAQGGYAVWVVDGKSVEKLAVK 1723

RESULT 6
Q9ZNB5 PRELIMINARY; PRT; 1223 AA.
ID Q9ZNB5 AC Q9ZNB5
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 130k-HMGD (Fragment).
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=99143166; PubMed=9988746;
RA Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
RT "Determination and characterization of the hemagglutinin-associated
RT short motifs found in Porphyromonas gingivalis multiple gene
RT products.";
RL J. Biol. Chem. 274:5012-5020(1999).
DR EMBL; AB019363; BAA34341.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.

```

```

FT NON TER 1 1
SQ SEQUENCE 1223 AA; 131542 MW; 00225CD2BA9F91B3 CRC64;
Query Match 70.7%; Score 6486; DB 2; Length 1223;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1211; Conservative 9; Mismatches 3; Indels 0; Gaps 0;
QY 510 NSWGEDYVWSGANAVFGVQTFEGTSMGSDATFLEDSTYNTVNSIMWAGNLAATHAGN 569
Db 1 NSWGEDYVWSGANAVFGVQTFEGTSMGSDATFLEDSTYNTVNSIMWAGNLAATHAGN 60
QY 570 IGNIITHIAHYWAEAHVLDGDSVMPYRAMPKTNITVTLPASLPQNASYSIOASAGSYVA 629
Db 61 IGNIITHIAHYWAEAHVLDGDSVMPYRAMPKTNITVTLPASLPQNASYSIOASAGSYVA 120
QY 630 ISKDGVLVGTGVANASGVATVMTKQITENGNDVVTITRSNLYPVTKIQVGPSPYQPV 699
Db 121 ISKDGVLVGTGVANASGVATVMTKQITENGNDVVTITRSNLYPVTKIQVGPSPYQPV 180
QY 690 SNLTATTQOKVTLKWEAPSAKKAEGSRVKRIGDGLFVTIIPANDVRANEAKVLLAADN 749
Db 181 SNLTATTQOKVTLKWEAPSAKKAEGSRVKRIGDGLFVTIIPANDVRANEAKVLLAADN 240
QY 750 VMGDNTGYQLLDADHNTFGSVIPATGFLTFGTASSNLYSANPEYLVPANADPVVTTQNI 809
Db 241 VMGDNTGYQLLDADHNTFGSVIPATGFLTFGTASSNLYSANPEYLVPANADPVVTTQNI 300
QY 810 IVTGQEVVPGVDYCIITNEPASGKMWIAGDGNQPARYDDFTFEAGKGYTFMTRA 869
Db 301 IVTGQEVVPGVDYCIITNEPASGKMWIAGDGNQPARYDDFTFEAGKGYTFMTRA 360
QY 870 GMGDGDMVEEDDPSASYTYTVYRDGTGKKEGLTATTFEEDGVAAAGHGYCUEVKYTAGV 929
Db 361 GMGDGDMVEEDDPSASYTYTVYRDGTGKKEGLTATTFEEDGVAAAGHGYCUEVKYTAGV 420
QY 930 SPKVCCKDVTVEGSNEFAPVQNLTGSSVGQKVLTKWDAPNGTNPNNPNPNTLSESF 989
Db 421 SPKVCCKDVTVEGSNEFAPVQNLTGSSVGQKVLTKWDAPNGTNPNNPNPNTLSESF 480
QY 990 ENGIPASWKTIDADGSHGWKPGNAPGIAGYNSNGCVYSESFLGGIGVLTPDNYLITPA 1049
Db 481 ENGIPASWKTIDADGSHGWKPGNAPGIAGYNSNGCVYSESFLGGIGVLTPDNYLITPA 540
QY 1050 LDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASFTNALBERTITAKGVRSFKAIR 1109
Db 541 LDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASFTNALBERTITAKGVRSFKAIR 600
QY 1110 GRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAADFTTFESSTHG 1169
Db 601 GRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAADFTTFESSTHG 660
QY 1170 EAPAEWTTIDADGGQGMCLSSGQLDWLTAHGGSNVVSFSGWNGMALPDNYLISKDVT 1229
Db 661 EAPAEWTTIDADGGQGMCLSSGQLDWLTAHGGSNVVSFSGWNGMALPDNYLISKDVT 720
QY 1230 GATKVKYVAVNDGPPGDHYAVMISKTGNAGDFTVVFETPTNGINKGARFGLSTEANG 1289
Db 721 GATKVKYVAVNDGPPGDHYAVMISKTGNAGDFTVVFETPTNGINKGARFGLSTEANG 780
QY 1290 AKPQSVWIERVTDLPAGTKYVAFRHHYNCSDLNIIILDDIOFTMGSGPTDPTDYTVYVRD 1349
Db 781 AKPQSVWIERVTDLPAGTKYVAFRHHYNCSDLNIIILDDIOFTMGSGPTDPTDYTVYVRD 840
QY 1350 TKIKEGLTETTFEEDGVATGNHGYCUEVKYTAGVSPKKCVDTVNSTQPNVQNLTAEOA 1409
Db 841 TKIKEGLTETTFEEDGVATGNHGYCUEVKYTAGVSPKKCVDTVNSTQPNVQNLTAEOA 900
QY 1410 PMSMDAILKNWAPASKRAEVLNEDFENGIPASWKITIDADGNGNWTTPPGSGSFAGHN 1469
Db 901 PMSMDAILKNWAPASKRAEVLNEDFENGIPASWKITIDADGNGNWTTPPGSGSFAGHN 960
QY 1470 SAICVSSASHINFEGQPNPDNYLTVPELSLPGGGLTFWVCAQDANYASEHYAVYASSTG 1529
Db 1470 SAICVSSASHINFEGQPNPDNYLTVPELSLPGGGLTFWVCAQDANYASEHYAVYASSTG 1529

```

Db 961 SAICVSASYINFRGPQNPNLYLTPPELPLPGLGCTLTFWCAODANYASEHYAVYASSTG 1020
Qy 1530 NDASNENALLEEVLTAKTVVTAPAEALRGCTWQKTVOLPAGTKYVAFRHFQCTDF 1589
Db 1021 NDASNENALLEEVLTAKTVVTAPAEALRGCTWQKTVOLPAGTKYVAFRHFQCTDF 1080
Qy 1590 FWNLDDVITSGNAPSYYTYIYRNNTQIASGVTTETTYRDPDLATGYTYGVKVVYPNGE 1649
Db 1081 FWNLDDVITSGNAPSYYTYIYRNNTQIASGVTTETTYRDPDLATGYTYGVKVVYPNGE 1140
Qy 1650 SAITATNITSLADVTAQPYTLTVVTKITVTCQGEAMIDYMNRRLLAAGRNTVVYTA 1709
Db 1141 SAITATNITSLADVTAQPYTLTVVTKITVTCQGEAMIDYMNRRLLAAGRNTVVYTA 1200
Qy 1710 QGHYAVWVVDGKSYVEKLVK 1732
Db 1201 QGHYAVWVVDGKSYVEKLVK 1223
RESULT 7
ID P96967 PRELIMINARY; PRT: 1358 AA.
AC P96967; 1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAGD.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA Han N., Jeline G., Whitlock J., Wojciechowski L., Progulski-Fox A.;
RT "Cloning, sequencing and characterization of hagD, a member of the
RT Harep multigene family in Porphyromonas gingivalis."
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U68468; AAB49691.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25_1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR SMART; SMO0060; FN3; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DE8D CRC64;
Query Match 67.1%; Score 6156.5; DB 2; Length 1358;
Best Local Similarity 85.0%; Pred. No. 2e-310;
Matches 1171; Conservative 59; Mismatches 118; Indels 29; Gaps 10;
Qy 366 MYEATWPKDKYLEKVLIIAGADYSWNSQVQPTIKYGMQYVYQNEHGYTDVNYLKAPY 425
Db 1 MYEATWPKDKYLEKVLIIAGADYSWNSQVQPTIKYGMQYVYQNEHGYTDVNYLKAPY 60
Qy 426 TGCYSHLNTGVSFANYTAGSETAWADPLLTSLQKALTNDKYLALGNCCITAAQEDYV 485
Db 61 TGCYSHLNTGVSFANYTAGSETAWADPLLTSLQKALTNDKYLALGNCCITAAQEDYV 120
Qy 486 QPCFGEVITRKEKAVAYIGSSPNSYMGEDYYSVGNANAVFGVQPTFEGTSMGSYDATF 545
Db 121 QPCFGEVITRKEKAVAYIGSSPNSYMGEDYYSVGNANAVFGVQPTFEGTSMGSYDATF 180

Qy 546 LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDSVMPYRAMPKNTY 605
Db 181 LEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVLGDSVMPYRAMPKNTY 240
Qy 606 TLPASLPQNASYSIQASAGSYVAISKDGYLYGTGVANASGVATVSMTKQITNGNYDVV 665
Db 241 TLPASLPQNASYSIQASAGSYVAISKDGYLYGTGVANASGVATVSMTKQITNGNYDVV 300
Qy 666 ITESNLYLPVTKIQVGEPSYQPVSNLTATTCQKQVTLKWEAPSAKKAEGSRVKRIGDG 725
Db 301 ITESNLYLPVTKIQVGEPSYQPVSNLTATTCQKQVTLKWEAPSAKKAEGSRVKRIGDG 360
Qy 726 LFVTIEPANDVRANEAKVLAADNVAGDNTGYQLLADAHNTFGSVIPATGPLTGTASS 785
Db 361 LFVTIEPANDVRANEAKVLAADNVAGDNTGYQLLADAHNTFGSVIPATGPLTGTASS 420
Qy 786 NLYSANFELYLVANADPVVTTQNIIVTQGEVVI PGVVDYCI TNPEPASGKMWIAGDGG 845
Db 421 NLYSANFELYLVANADPVVTTQNIIVTQGEVVI PGVVDYCI TNPEPASGKMWIAGDGG 480
Qy 846 NOPARYDDTFEAGKVTFTMRAGMGDTMEVEDDSDPASYYTVVTRDGTIKELGTAT 905
Db 481 NOPARYDDTFEAGKVTFTMRAGMGDTMEVEDDSDPASYYTVVTRDGTIKELGTAT 540
Qy 906 TFEEDGVAAGNHEVCVEVKYTAGVSPKVKDVTVEGSNEEPAPVQNLTSVGQKVLKWD 965
Db 541 TFEEDGVAAGNHEVCVEVKYTAGVSPKVKDVTVEGSNEEPAPVQNLTSVGQKVLKWD 600
Qy 966 APNGTNPNNPNPGTTLSESFENGIPASWKTIDADGGHGWKPNAGIAGINSNC 1025
Db 601 APNGTNPNNPNPGTTLSESFENGIPASWKTIDADGGHGWKPNAGIAGINSNC 659
Qy 1026 VYSESFLGGIGVLTDPNLYLITPALDLPNGKGLTFWCAODANYASEHYAVYASSTGND 1085
Db 660 VYSESFLGGIGVLTDPNLYLITPALDLPNGKGLTFWCAODANYASEHYAVYASSTGND 719
Qy 1086 SNFTNALLEETITAKGVRSKPAIRIGTQWTKVTDLPAGTKYVAFRHFQSTDMFYIDL 1145
Db 720 SNFTNALLEETITAKGVRSKPAIRIGTQWTKVTDLPAGTKYVAFRHFQSTDMFYIDL 779
Qy 1146 DEVEIKANGRADFTETTFESTHGEAPAEWTTIDADGGQGLCLSSGQDLMTAHGGSN 1205
Db 780 DEVEIKANGRADFTETTFESTHGEAPAEWTTIDADGGQGLCLSSGQDLMTAHGGSN 839
Qy 1206 VVSFSWNGMALPNPNLYISKDVTGATKVKYVAVNDGFFPDHVAVMISKTGNAGFTV 1265
Db 840 VVSFSWNGMALPNPNLYISKDVTGATKVKYVAVNDGFFPDHVAVMISKTGNAGFTV 899
Qy 1266 VFETPNNGKNGARFGLSTEANGAKPQSVIERTVDLPAGTKYVAFRHNCSDLNLYL 1325
Db 900 VFETPNNGKNGARFGLSTEANGAKPQSVIERTVDLPAGTKYVAFRHNCSDLNLYL 959
Qy 1326 DDQFTMGGSPTDVTYTVTRDGTIKELGTTETFEEDGVATGNHEVCVEVKYTAGVSP 1385
Db 960 DDQFTMGGSPTDVTYTVTRDGTIKELGTTETFEEDGVATGNHEVCVEVKYTAGVSP 1019
Qy 1386 KKCVDVTNSTQPNVQNLTAEQAPNSMDAILKWNAPASKRAEVLNEDFE-NGIPASWKT 1444
Db 1020 KVCNVVNTINFPNPKNLKAQ--PDGDDVVLKWEAPSGKRGLLNEDFEEDGALPTGWA 1077
Qy 1445 IDADGGGNV--TTTPPPGG-----SFPAGHNSAICVSSASHINFECPQPNLYLTPEL 1497
Db 1078 LDADGGGNNDITLNEFTRGERHVLSPLRASNVAISYSLLOQOEYLPPLTNFLTPKV 1137
Qy 1498 SLPGGGLTFWCAQD--ANVASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPAEI 1556
Db 1138 E--GAKKITVKGSPGLPQWSHDHYALCISKSTAAADF-----EVIETWTYQCGA 1189
Qy 1557 RGTRAQCTVTKVQVPAQTKYVAFRHFQCTDFWNLDDWIT-SGNAPSYTYTYRNN 1615
Db 1190 NLTRK-----DLPAQTKYVAFRHNCTDVLGIMIDDVITGEGEGSYTYTVYRDG 1241
Qy 1616 TQIASGVTTETTYRDPDLATGFYTVGVKVPNGESAETATLNTSLADVTAQPYTLTV 1675

```

Db      1242 TKIQSLTETTYRDAGMSAQSHYCVVEKYAAGVSPKVCVDYIPDGVADVTAKPXTLTV 1301
QY      1676 VGTITVTTCQGEAMIDMNGRRLAAGRTVVYTAQGGHYAMVVDGKSYVEKLAIV 1732
Db      1302 VGTITVTTCQGEAMIDMNGRRLAAGRTVVYTAQGGHYAMVVDGKSYVEKLAIV 1358

RESULT 8
Q51839 PRELIMINARY; PRT; 1706 AA.
AC Q51839; Q51840;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arginine-specific thiol protease precursor.
GN PRT.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=95160709; PubMed=7857299;
RA Kirsbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA Reynolds E.C.;
RT "Complete nucleotide sequence of a gene prtr of Porphyromonas
RT gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT specific thiol endopeptidase domain and a haemagglutinin domain.";
RL Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=96311339; PubMed=8713096;
RA Slakeski N., Cleal S.M., Reynolds E.C.;
RT "Characterization of a Porphyromonas gingivalis gene prtr that encodes
RT an arginine-specific thiol proteinase and multiple adhesins.";
RL Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Reynolds E.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Slakeski N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L26341; AAC18876.1; -.
DR HSSP; P95493; 1CVR.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW Protease; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 228 1706 ARGinine-SPECIFIC THIOL PROTEASE.
SQ SEQUENCE 1706 AA; 185626 MW; E8BDF07C9813B844 CRC64;

```

Query Match 46.5%; Score 4272; DB 2; Length 1706;
 Best Local Similarity 50.9%; Pred. No. 1.4e-212;
 Matches 932; Conservative 204; Mismatches 456; Indels 240; Gaps 47;

```

QY      6 LLIAASLLGVGLYAQSAKIKLDAPTRTCTTNSFKQDPASFSFNEVELTKVETKGTFA 65
Db      10 IALCSLLGLGMFAAQTELGRLNPNVRLLESTQOSTK--VQFRMDNLKFTVOTPKGI-- 65
QY      66 SVSISGAFT-----GEVGSPEVPAVRKLTAVPGATPVVRKVS--FTEOVYSLNQXG 116
Db      66 -----GOVPTYTEGVNLSEKGMPTPIILSRSLAVSDTREMKVEVVSKEIEKNVL---- 116
QY      117 SEKLPHQPSMSKSDDEPKVPFYVNAAYARKFGVQELTQVEMGLTGMGVRAALATINP 176
Db      117 ---IAPSKGIMRNEDPKKIPYV--GKYSQNKFFPGELATLDDPEILRDVRGVVNFAP 172
QY      177 VQYDVVANQLKYRNNEI---EVSFGQDADEVATQRLYDASFPSPETAKVQFNRDVYTD 233
Db      173 LQINPVTKLRIYTEITVAVSETSQGNILNKGTFAG-----FEDTYKRMF-----WN 222
QY      234 HGDLYNTPV-----RMLVVGAKFKKALKPWLTKWAKQGFYLDVHYTDAEAVGTTNASI 287
Db      223 YEPGRYTPVEEKQNGRMIVIVAKYEGDIKDFVDMKNQGLTEVKVAEDIASPVTANAI 282
QY      288 KAFIHKKY---NDGLAASAAPVFLALVGDTDVISE--KGKTKVKVTLDTYSAVDGYF 341
Db      283 QQFVKQYEKEGND-----LTYVLLIGHDKDIPAKITPGIKSDQV---YQIVGNHY 332
QY      342 PEMYTFRMSASSPEELTNIIDKVLAKYKATMPDKSVLEKVLIIAGADYSWNSVQGOPTIK 401
Db      333 NEVFTGRSCESKEDLKTQIDRTIHYERNITTEDKWLQALCLASABGGSADNGSDLIQ 392
QY      402 Y-GMOYVYNQEHGYTDVNYLNKAPYTGCV-----SHLNTGVSFANYTAHSETAW 450
Db      393 HENVIANLLTQYGTKIILK-----CYDPGVTPKNIIDAENFGGISLANYTGHSETAW 444
QY      451 ADPLLTTSQKALKTNKDYFLAIGNCCITAAQPDYVQPCFGEVITRV-----KEKGAYAYIG 506
Db      445 GTSHEFTTHVKQLTNSNQLPFIFDVACVNGDPLFSMPCFAEALMRAQDKGKGTGTVAIIA 504
QY      507 SSPNSVWGEDYVWSVCANAVFGVQPTFEGTSMGSDATFLEDSYNTVNSIMWAGNLAATH 566
Db      505 SLINQSW-----ASPMRG-----QDEMNEI-----LCEKH 529
QY      567 AGNI-----GNITHIGAHYWEAYH-----VLGDGVMPYRAMPKNTYTLPASL 611
Db      530 PNNIKRTFGVTMGFMFAMVEKYKDKGEKWLDTWTFGDPSSLVRLTLVTKMQVTAQAQI 589
QY      612 PQNQASYSIQAS-AGSYVAISKDGLVYGTGVANASCVATVSMTKQITENGNDVIVTRSN 670
Db      590 NLTDASVNVSCDYNAGIAIATISANGKMFSAVVE-NGTATINLT-GLTNBSTLTTLVVGYN 647
QY      671 YLPVIKIQV-GEPSYPQVSNLTATTQOKVYTLKWEAPSAK---KAEGSREVKRIQDGL 726
Db      648 KETVIKTINGEPNPYPQVSNLTATTQOKVYTLKWDAPSTKTNATNTARSVDGIRELV 707
QY      727 FVTISPAND-VRANEAKVVLAAADNVGMDNTGYOFLDADHNTFGSVIPA-TGELFTG-TA 783
Db      708 LLSVSDAPELLRSQAEIVLEAHDVWMDGSGYQILLADHDQYGVQVPSDTHLTPNCVS 767
QY      784 SSNLYSANEFLVAPNADPVTQNLIVTGQGEVVIIPGVGYDVCITNPPSPASCKWIAQD 843
Db      768 PANLF-APPEYTVPENADPSCPTNMIMDGTASVNIIPAGTYDFAIAAPQ-ANAKIWIAGQ 825
QY      844 GGNQPARYDDFTFEAGKKYTFMTRAGMGDGMEDVEDDPSASYTVTVVYRDGTCKIEGLT 903
Db      826 G---PTKEDDYVFEAGKKYHFLMKWMSGDGLFELISEGGSDYTYTVVYRDGTCKIEGLT 882
QY      904 ATTFEDGVAAGNHCVVEVKYTAGVSPKVCXDVTVGSENEFAPVQNLTGSSVGQKVTUK 963
Db      883 ATTFEDGVATGNHCVVEVKYTAGVSPKVCXDVTVGSENEFAPVQNLTGSAVGQKVTUK 942
QY      964 WDAPNGTNPENPNPNPG---TTLSSEFENGIPASWKTIDADGGHGWKPNAGIACY 1020
Db      943 WDAPNGTNPENPNPNPNPGTTTLESFENGIPASWKTIDADGGHGWKPNAGIACY 1002

```

QY 1021 NSNGCVYSESGLGIGVLTDPNYLITPALDPLNGGKLTFFWCAQDANYASEHVAIVYASS 1080
Db |||||
QY 1003 NSNGCVYSESGLGIGVLTDPNYLITPALDPLNGGKLTFFWCAQDANYASEHVAIVYASS 1062
Db |||||
QY 1081 TGNDSNTFNLLETTITAKVRSPEAKIRGRIQGTWRQKTVDLDPAGTKYVAFRHFQSTDM 1140
Db |||||
QY 1063 TGNDSNTFNLLETTITAKVRSPEAKIRGRIQGTWRQKTVDLDPAGTKYVAFRHFQSTDM 1122
Db |||||
QY 1141 FYDLDEVEIKANGKRADETFEFTSSSTHGEAPAEWTTIDADGGGMLCLSSGQDMLTA 1200
Db |||||
QY 1123 FYDLDEVEIKANGKRADETFEFTSSSTHGEAPAEWTTIDADGGGMLCLSSGQDMLTA 1182
Db |||||
QY 1201 HGGSNVVSFSWNGMALNPDNYLISKVYATGATKVKYVAVNDGPFDDHYAVIMISKTGTNA 1260
Db |||||
QY 1183 HGGTNVVSFSWNGMALNPDNYLISKVYATGATKVKYVAVNDGPFDDHYAVIMISKTGTNA 1242
Db |||||
QY 1261 GDFTVVFETENGKNGARGLSTEANGAKPQSWIERTVDLDPAGTKYVAFRHYNCSDL 1320
Db |||||
QY 1243 GDFTVVFETENGKNGARGLSTEANGAKPQSWIERTVDLDPAGTKYVAFRHYNCSDL 1302
Db |||||
QY 1321 NVILLDDIQFTMGSPPTDYTYTVYRDGTIKI KEGLTETTFEEDGVATGNHEYCVVKYT 1380
Db |||||
QY 1303 NVILLDDIQFTMGSPPTDYTYTVYRDGTIKI KEGLTETTFEEDGVATGNHEYCVVKYT 1362
Db |||||
QY 1381 AGVSPKCVDVVNSTQNPVONLTAEQAPNSMDAILKWNAPASKRAE----- 1428
Db |||||
QY 1363 AGVSPKCVDVVNSTQNPVONLTAEQAPNSMDAILKWNAPASKRAE----- 1420
Db |||||
QY 1429 -----VLNEDFENGIPASWKITDADGGNNWTTTPPPGSSFAGHNSA 1471
Db |||||
QY 1421 FVTIEPANDVRANEAKVLAADNVWGDNTGYQL-LDADHNTFGSVIPATGFTGTASS 1479
Db |||||
QY 1472 ICVSASASHINEG--PONPD-----NYLVT--PELSPPGGGTLTFWYCAODANYASEHY 1521
Db |||||
QY 1480 DLYSA-----NFSLEIPANADPVITQNIIVTQGEVVIPIGG--VYDCIINPEPASKM 1532
Db |||||
QY 1522 AVYASSTGNDASNFANALLEVLFAKTVVTAPEAIRGTAQGTWYQTVQVLPAGTKYV-A 1580
Db |||||
QY 1533 WI-AGDGNQPARYDDTFE-----AGKKYTF 1559
Db |||||
QY 1581 FRHFGCTDFWNLDDVITSGNAPSVTYTYRNTQIASGVTTTTRDPLATGFTYTG 1640
Db |||||
QY 1560 MRRAGMDG-----TDMVEDDSPASYTYTVYRGDTIKI KEGLTETTYRDAGMSAQSHCYC 1614
Db |||||
QY 1641 VKVYVNGESAETATNLITSLADVTAQKPYTLTVVGKTIITVTCQGEAMIDMNGRLAA 1700
Db |||||
QY 1615 VEVKYTAGVSPKVDYIPDGVADVTAQKPYTLTVVGKTIITVTCQGEAMIDMNGRLAA 1674
Db |||||
QY 1701 GRNTVVYTAQGGHYAVMVVDGKSYVEKLAVK 1732
Db |||||
QY 1675 GRNTVVYTAQGGHYAVMVVDGKSYVEKLAVK 1706
Db |||||

RESULT 9

Q51816 PRELIMINARY; PRT; 1704 AA.
ID Q51816
AC Q51816;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arg-gingipain-1 proteinase.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95138080; PubMed=7836351;
RA Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
RA Travis J., Barr P.J.;
RT "Molecular cloning and structural characterization of the Arg-
gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
proteinase-adhesin polyprotein.";

J. Biol. Chem. 270:1007-1010 (1995).
EMBL; U15282; AAA69539.1; -.
PIR; A55426; A55426.
HSPF; P95493; 1CVR.
MEROPS; C25.001; -.
GO; GO:0005524; P:ATP binding; IEA.
GO; GO:0008234; P:cysteine-type peptidase activity; IEA.
GO; GO:0003910; P:DNA ligase (AUP) activity; IEA.
GO; GO:0006310; P:DNA recombination; IEA.
GO; GO:0006281; P:DNA repair; IEA.
GO; GO:0006260; P:DNA replication; IEA.
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
InterPro; IPR000977; DNA ligase.
InterPro; IPR001110; Ig-Like.
InterPro; IPR001769; Peptidase_C25.
InterPro; IPR005536; Peptidase_C25-C.
Pfam; PF01364; Peptidase_C25; 1.
Pfam; PF03785; Peptidase_C25; 1.
PROSITE; PS00697; DNA_LIGASE_A1; 1.
CHAIN 228 719 MATURE 50-KDA CYSTEINE PROTEINASE
FT GINGIPAIN
ET
SQ SEQUENCE 1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;
Query Match 46.5%; Score 4270; DB 2; Length 1704;
Best Local Similarity 50.9%; Pred. No. 1.8e-212;
Matches 932; Conservative 204; Mismatches 456; Indels 238; Gaps 47;
QY 6 LLTAASLLGLVLAQSAKIKLDAPTRTTCTTNNSTKFPDASFSFNEVELTKVETKGTFA 65
Db |||||
QY 10 IALCSSLLGMAFAQQTLEGRNPNVRLLESTQOSVTK--VQFRMDNLKFTVQTPKGI-- 65
Db |||||
QY 66 SVSIPGAPPT-----GEVSPVPAVRKLIAPVPGATPVVRKS--FTEQVYSLNQYQ 116
Db |||||
QY 66 -----GOVPTTEGVNLSEKGMTPLPILSRSLAVSDTREMKVEVSSKFIKKNVL---- 116
Db |||||
QY 117 SEKLMPHQPSMSKSDDEKVPFVYNAAYARKGVGOELTQVEMLGPMRGVRIAAALINP 176
Db |||||
QY 117 ---IAPSKGMIRNEDPKIPYV--GKSYSONKFFPGCEIATLDDPFLRDVRGQVNFAP 172
Db |||||
QY 177 VOYDVVANQLKVRNNIEI---EVSFGADEVAQRLYDASFSFYFETAYKQLFNRDVTYD 233
Db |||||
QY 173 LQNPVTKTLRIYETITVAVSETSEQKNTLNKKGTFAG-----FETYKRMF-----MN 222
Db |||||
QY 234 HGDLNTPV-----RMLVVAGAKFKEALKPMLTWKAQGFYLDVHYTDAEAVGTNNAI 287
Db |||||
QY 223 YEGRYTPVEEKQNGRMIVIVAKYEGDIKDFVDMKNQGRGLRTEVKVAEDIASPVITANAI 282
Db |||||
QY 288 KAFIHKY-----NDGLAASAAPFLALVGDVDSIGE--KGKTKKVTDLIYSAVDGDF 341
Db |||||
QY 283 QQFVKQYEKEGND-----LTYVLLVGDHDKIPAKITPGIKSDQV---YQQIVGNHDY 332
Db |||||
QY 342 PEMYTPRMSASSPEELNIIIDKVLWYEKATMPDKSYLEKVLIIAGADYSNNSQVQPTIK 401
Db |||||
QY 333 NEVFIGRFSCKEDLTKQIDRIHVERNTTETDKLQALCIASAGGGSANDGESDIQ 392
Db |||||
QY 402 Y-GMQYYINQEHGYTDVNYLKAPYTGCTY-----SHLNTGVSFANYTAHGSSTAW 450
Db |||||
QY 393 HENVIANLLTQYGVTKIILK-----CYDPGVTPKNIIDAFNGGISLVNVTGHSSTAW 444
Db |||||
QY 451 ADPELLTTSQKALTKDKKYFLATGNCCIITAQFDYVQPCFGEVITRV-----KEGAYAYIG 506
Db |||||
QY 445 GTSHFGTTHVKQLTNSGNQLPFIIDVACVNGDGLFSMPCFABALMRAQKQKPTGTVAIIA 504
Db |||||
QY 507 SSPNSWYGEDYVSWGVANAVFGVQPTFEGTSMGSDATFLEDISYNTVNSIMWAGNLAATH 566
Db |||||
QY 505 STINQSW-----ASPMRG-----QDEMNEI-----ICEKH 529
Db |||||
QY 567 AGNI-----GNITHIGAHYWEAYH-----VLGDGSMVPMYRAMPKNTNTYTLPASL 611
Db |||||
QY 530 PNNIKRTFGVTVNGWFAMVEKYKDKGKMLDNTWTFGDPSSLVRLVPTKMQVTAPAQI 589
Db |||||
QY 612 PQNQASYSIQAS--AGSYVAISKGVLYGTGVANASGVATVSMTKQITENGNYDVVITRSN 670
Db |||||

Db 590 NLTDASVNSCDYNGALATISANGKMPGSAVVE-NGTATINLT-GLTNESTLTLTVVGYN 647
QY 671 YLPVILQIQV-GRPSXPQPSNLTATTOGQKVLKWEAPSAK--KAGSREKRTGDGL 726
Db 648 KETVIKTIINTNGBNPYQPSNLTATTOGQKVLKWDAPSTKTATNTARSVDGIRELV 707
QY 727 FVTIEPAND-VRAEAKVLAADNVDGNTGYOFLDADHNTFGSVIPA-TGPLFTG-TA 783
Db 708 LLSVSDAPELLRSGQAEIVLEADHVNDGSGYQLLDADHDQGVQVIPSOTHTLWNCV 767
QY 784 SSNLYSANFYLVPANADPVVTTQNIIVTQGEVVIPIGGVYDICTNPEPASGMKIAGD 843
Db 768 PANLF-APFEYTVENADPSCSPNTIMDGTASVNPAGTYDEAIAPO-ANAKIMIAGQ 825
QY 844 GSNQPARYDOFTFAGKKTFTMRAGMGDTGMEVEDDSPASYTYTVYEDGKIKEGLT 903
Db 826 G--PTKEDDYFEAGKKHFLMKMGSGDGTLTISEGGSYTYTVYEDGKIKEGLT 882
QY 904 ATTFEEDGVAAGNHEVCVEVKYTAGVSPKCVKDVTEGSGNEFAPVQNLTGSSVGQKVTLK 963
Db 883 ATTFEEDGVAAGNHEVCVEVKYTAGVSPKCVKDVTEGSGNEFAPVQNLTGSAVGQKVTLK 942
QY 964 WDPNGTNPNNPNPNPG--TTLSSESPENGIPASWKTIIDADGCHGKWPKNAPGCIAGYNS 1022
Db 943 WDPNGTNPNNPNPNPGTTLSSESPENGIPASWKTIIDADGCHGKWPKNAPGCIAGYNS 1002
QY 1023 NGCVYSESGLGIGVLTDPNLYLTALDLPNGKLTFTWCAQDANYASHYAVYASSTG 1082
Db 1003 NGCVYSESGLGIGVLTDPNLYLTALDLPNGKLTFTWCAQDANYASHYAVYASSTG 1062
QY 1083 NDASNFTNALLEETITAKGVRSKPAIRIGRIQGTWRQKTVDLPAGTKVAFRHFQSTDMFY 1142
Db 1063 NDASNFTNALLEETITAKGVRSKPAIRIGRIQGTWRQKTVDLPAGTKVAFRHFQSTDMFY 1122
QY 1143 IDLDEVEIKANGRADPTTFESTHGEAPAEWTTIDADGGQWLCLSSGQDLWLTAGH 1202
Db 1123 IDLDEVEIKANGRADPTTFESTHGEAPAEWTTIDADGGQWLCLSSGQDLWLTAGH 1182
QY 1203 GSNVVSFSNGMALPNPNLYLSKQVTGATKVKYVAVNDGFGDHYAVMISKTGTNAGD 1262
Db 1183 GTNVVASFSGMALPNPNLYLSKQVTGATKVKYVAVNDGFGDHYAVMISKTGTNAGD 1242
QY 1263 FTVVFEETPNNGKGFGLSTEANGAKPQSVMIERTVDLPAGTKVAFRHYNCSDLNY 1322
Db 1243 FTVVFEETPNNGKGFGLSTEANGAKPQSVMIERTVDLPAGTKVAFRHYNCSDLNY 1302
QY 1323 ILDDOFTQMGSGPTPDYTYTVYRDGTKIKEGLTTFEDGVATGNHHCYVEVKYTAG 1382
Db 1303 ILDDOFTQMGSGPTPDYTYTVYRDGTKIKEGLTTFEDGVATGNHHCYVEVKYTAG 1362
QY 1383 VSPKCKVDVTNVTQFNPNVQNLTAEOAPNSMDALLKNAPASKRAE----- 1428
Db 1363 VSPKCKVDVTNVTQFNPNVQNLTAEOAPNSMDALLKNAPASKRAE----- 1408
QY 1429 -----VLNEDFENGIPASWKTIIDADGGNNWTTTPPPGGSSFAGHNSAIC 1473
Db 1421 TIEPANDVRAEAKVLAADNVDGNTGYOFL-LDADHNTFGSVIPATGFLFTGTASSNL 1479
QY 1474 VSSASHINFG--PONPD-----NYLVT--PELSLPGCGTLTFWCAQDANYASHYAV 1523
Db 1480 YSA-----NFYLLIPANADPVVTTQNIIVTQGEVVIPEG--VYDYCITNPEPASGRMWI 1532
QY 1524 YASSTGNDASNFANALLEEVLTAKTVVVTAPEAIRGTRAQGTWQKTVQLPAGTKYV-APR 1582
Db 1533 -AGGGNQPARYDDFTFE-----AGKTYFTMR 1559
QY 1583 HFGCTDPFNLDDVITSGNAPSYYTYIYRNNTQIASGVTTETTYRDPDLTGFYTYGVK 1642
Db 1560 RAGMGDG-----TDMEVEDDSPASYTYTVYRDGTKIKEGLTETTYRDAGNSAQSHCYVE 1614
QY 1643 VTPNGESALETATLNTLSADVTAQKPYTLTVVVGKTIITVTCQGEAMIDMNGRRLAAGR 1702
Db 1615 VKYAAGVSPKVCVDYIPDGVDADVTAQKPYTLTVVVGKTIITVTCQGEAMIDMNGRRLAAGR 1674

QY 1703 NTVVYTAQGHYAVMVVDGSKSYVEKLVAK 1732
Db 1675 NTVVYTAQGHYAVMVVDGSKSYVEKLVAK 1704

RESULT 10

QY 1838 PRELIMINARY; PRT; 1706 AA.
ID Q51838
AC Q51838;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Protease precursor.
GN PAPR1.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
CX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=96071894; PubMed=7591131;
RA Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
RT "Characterization, genetic analysis, and expression of a protease
antigen (Prp81) of Porphyromonas gingivalis W50.";
RL Infect. Immun. 63:4744-4754 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
RT "The prp81 and the prp82 arginine-specific protease genes of
Porphyromonas gingivalis W50 produce five biochemically distinct
enzymes.";
RL Mol. Microbiol. 23:0-0 (1997).
DR EMBL; X82680; CAA57997.1; --
DR HSP; P95493; 1CVR.
DR MEROPS; C25_001; --
DR GO; GO:0005524; F-ATP binding; IEA.
DR GO; GO:0008234; F-cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F-DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P-DNA recombination; IEA.
DR GO; GO:0006281; P-DNA repair; IEA.
DR GO; GO:0006260; P-DNA replication; IEA.
DR GO; GO:0006508; P-proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW SIGNAL.
FT SIGNAL. 1 23 POTENTIAL.
FT CHAIN 228 719 ALPHA-PROTEASE.
FT CHAIN 720 1262 BETA-ADHESIN.
SQ SEQUENCE 1706 AA; 185705 MW; 0556DCD87FDA8CDD CRC64;

Query Match 46.4%; Score 4256; DB 2; Length 1706;

Best Local Similarity 50.8%; Pred. No. 9.7e-212;

Matches 928; Conservative 206; Mismatches 465; Indels 226; Gaps 45;

QY 6 LLIAASLLGVGLYAQSAKIKLDAPTTTCTNNSPKQDFDSFNEVELTKVETKGTGA 65
Db 10 IAUSSLLGMAFAQQTGELGRNPVLLSTQOSTQVTK--VQFRMDLKFTEVQTPKGMQ 67
QY 66 SVSIPGAFPTGEVGSPEVPAVKLIAVPVGCATPVVRVKS--FTEQVYSLNQYSEKIMPH 123
Db 68 VPITYTEGNLSKGMETLILSRSLAVSTREMKVEVSKFIEKNVL-----IAPS 120
QY 124 QFSMSKDDPEKVPFYVYNAAYARKGVQGLTQVEMLTMRGVRIAALTINPVQDVVA 183
Db 121 KGMIMENEDPKIPYVY--GKSYSONKFFPGELATLDDPFLDRVQGVVNFAPLQNPVT 179

184 NQKVRNNIEI---EVSFGADEVATQRLYDASPSPEETAYKQLFNRDVTYDHGDLNLT 240
180 KTLRIYETITVAVSETSQGKNILNKKGTAG-----FEDYTKRMF-----MYEPGRYT 229
241 PV-----RMLVAGAKFEKALPMLTWKAKGFYLDVHYTHDEAEVGTNTASIKAFIHK 294
230 PVEEKQNGRMIVIVAKKYEKDIKDFVDMKNQGLRTERVKVAEDIASPVTAIAIQFVKQE 289
295 Y-----NDGLAASAPVFLALVGDITVIGSE--KGKTKKVTDLIYSVAVDGYDFPEMYTFR 348
290 YEKEGND-----LTYVLLGLDGHKDIKPAKTPGIKSDQV---YQIIVGNDHYNEVFGR 339
349 MSASSPELTIIDKLVMEYKATPDKSYLEKVLIIAGADYSWNSOVGPQTIK-GMQYY 407
340 FSCSKEDLKQIDRTHIYERNITFEDKWLQALCIASAGSPADNGESDIQHENVIAN 399
408 YNQBHGVTDVNYLKAFTVGCY-----SHLNTGVSFANYTAHGETAWADPLLT 457
400 LLTQVGYTKIHK-----CYDPGVTPKNIIDAFNGGIGISLANYTGHGSETAWGTSHEGT 451
458 SOLKALTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV-----KEKGAYAVIGSPNSYW 513
452 THVKQLTNSNQLPFTFDVACVNGDFLFSMPCFAEALMPRAQDKGPTGTVAIIASTINGSW 511
514 CEDYVWSYGANAVFGVQPTFEETSGMSYDATFLEDSYNTVNSIMWAGNLAATHAGNI--- 570
512 -----ASPMRG-----QDEMNEI-----LCEKHPNNIKRT 536
571 -GNTHIGHAYWEAYH-----VLGDSVMPYRAMPKNTYTLPLASLPQNASY 618
537 FGGVTMNGMFAMVEKYKDKGEMALDTWTVFGDPSLLVRLTLPVKMQVTPAQINLTDAV 596
619 SIQAS-AGSYVAISKDGLYGTGVANASGATVSMKTQITENGVDVVTIRSNVLPVKQ 677
597 NVSCDYNRAITISANGKMFSAVVE-NGTATINLT-GLTNESTLTLTVVGVNKETVIKT 654
678 IQV-GEPSYQFVSNLTATTQKQVTLKWEAPSAK---KASGSRVRKRGIDGLFTIETPA 733
655 INTNGEPNYPQVSNLTATTQKQVTLKWDAPSTKTNNATTNARSVDGIRELVLLSVSDA 714
734 ND-VRANEAKVLAADNVGNTGQFLLDADHNTFGSVIPA-TGPLFTG-TASSNLYSA 790
715 PELLRSQAETVLEAHDVMDGSGYQIILLADHDQGVQIFSDTHLWPNCSVFNALF-A 773
791 NFEYLVANADPVVTTQNIITVQGEVVI PGVVDYCLITNPEPAGKMWIAGDGNQPAR 850
774 PFETVVENADPSCPTNIMIDGTASVNI PAGTYDFAIAAQ-ANAKIWIAGQG---PTK 829
851 YDDFTFEAGKKYFTMRRAAGMDGTDMEVEDDSPASYTYTYVRDGTIKIKEGLTATTFEED 910
830 EDDYVFEAGKKYHFLMKMGSGDGTETATISEGGGSDYTYTYVRDGTIKIKEGLTATTFEED 889
911 GVAAGNHCYCEVKYTAGVSPKCVKDVTVGSENEFAPVQNLTGSSVGOKVTLKWDAPNGT 970
890 GVAAGNHCYCEVKYTAGVSPKCVKDVTVGSENEFAPVQNLTGSAVGOKVTLKWDAPNGT 949
971 PNPENPNPNPFG---TTLSESPENGIPASWKTIDADGDGHWKPGNAPAGIAGYNSGCVY 1027
950 PNPENPNPNPNPFGTTLISESPENGIPASWKTIDADGDGHWKPGNAPAGIAGYNSGCVY 1009
1028 SESFGLGGIGVLTDPNLYITPALDIPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASN 1087
1010 SESFGLGGIGVLTDPNLYITPALDIPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASN 1069
1088 FTNALLEETITAKGVRSKPAKTRGRTQGTWRQKTVDLPAKTKYVAFRRHQSTDMFIDIDE 1147
1070 FTNALLEETITAKGVRSKPAKTRGRTQGTWRQKTVDLPAKTKYVAFRRHQSTDMFIDIDE 1129
1148 VEIKANGKRAOPTFETFEESSTHGEAPAEWTTIDADGDGQWCLLSSGQLDWLTAHGGSNV 1207
1130 VEIKANGKRAOPTFETFEESSTHGEATAEWTTIDADGDGQWCLLSSGQLDWLTAHGGSNV 1189

QY 1208 SSFSWNGMALPNPNYLISKDVTKYKVVYVNDGPDGHYAVIMISKTGTCNAGDFTVWF 1267
Db 1190 SSFSWNGMALPNPNYLISKDVTKYKVVYVNDGPDGHYAVIMISKTGTCNAGDFTVWF 1249
QY 1268 EETPENGINKGAREGLSTEANGAKPQSVWIERTVDLTPAGTKYVAFRRHVCSDNLNILLDD 1327
Db 1250 EETPENGINKGAREGLSTEANGAKPQSVWIERTVDLTPAGTKYVAFRRHVCSDNLNILLDD 1309
QY 1328 IQFTMGSGPPTDYTYTYVRDGTIKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKK 1387
Db 1310 IQFTMGSGPPTDYTYTYVRDGTIKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKK 1369
QY 1388 CVDVTNASTQFNPNVONLTAEQAQNSMDAILKNWAPASKRAE----- 1428
Db 1370 CVDVTNASTQFNPNVONLTAEQAQNSMDAILKNWAPASKRAE----- 1427
QY 1429 -----VLNEDFENGIPASWKTIDADGDGNWNTTTPPPGSSPAGHNSAICVSSAS 1478
Db 1428 NDVRANEAKVLAADNVGNTGQFL-LDADHNTFGSVIPATGFLFTGTASSDLISA-- 1484
QY 1479 HINEPEG--PQNPD-----NYLVT--PELSLPGGGTILTFWVCAQDANYASEHYAVYASST 1528
Db 1485 --NFEYLIPANADPVVTTQNIITVQGEVVIPEG---VYDYCIITNPEPASGKMI-AGDG 1538
QY 1529 GNDASNEFANALLEEVLTAKTVTWTAPEAIRGTRAQGTWYQKTVQLPAGTKV-APRHTGCT 1587
Db 1539 GNQPARYDDFTFE-----AGKKYFTTMRAGM 1566
QY 1588 DFFWINLDDVVITSGNAPSYYTYTYRNNNTQIASGVTETTRDPLATGTYTYGVKVVYPN 1647
Db 1567 DG-----TDMVEVDDSPASYTYTYVRDGTIKIKEGLTETTRDAGMSAQSHCYCVVEKYTA 1621
QY 1648 GESALETATLNTISLADVTAKPVTLVGKTIITVTCOGEMIVDMNGRLAAGNTVWY 1707
Db 1622 GVSFKVCVDYIDGVAADVTAQKPYTLTVGKTIITVTCOGEMIVDMNGRLAAGNTVWY 1681
QY 1708 TAQGGHYAVMVVVDGKSYVEKLAVK 1732
Db 1682 TAQGGHYAVMVVVDGKSYVEKLAVK 1706
RESULT 11
P72196 PRELIMINARY; PRT; 1097 AA.
AC P72196;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TonB-linked adhesin precursor.
GN TLA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_taxid=837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97386416; PubMed=9244265;
RA Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA Curtis M.A.;
RT "The tia gene of Porphyromonas gingivalis W50: a homologue of the
RT arginine-specific protease precursor (PrpRI) which shares sequence
RT similarity to TonB-linked receptors.";
RL J. Bacteriol. 179:4778-4788(1997).
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Y07618; CAA68897.1; -;
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.

| | |
|--|--|
| DR | GO: 0006810; P: transport; IEA. |
| DR | InterPro; IPR000977; DNA_ligase. |
| DR | InterPro; IPR000531; TonB_box. |
| DR | Pfam; PF00593; TonB_dep_Rec; 1. |
| DR | PROSITE; PS00697; DNA_LIGASE_A1; 1. |
| KW | Membrane; Outer membrane; Receptor; Signal; TonB_box. |
| FT | SIGNAL 1 53 |
| SQ | SEQUENCE 1097 AA; 118731 MW; 73BBA37B421F89 CRC64; |
| Query Match | |
| Best Local Similarity 46.3%; Score 4249.5; DB 2; Length 1097; | |
| Matches 824; Conservative 31; Mismatches 44; Indels 91; Gaps 12; | |
| Qy | 796 VPANADPVVT--TONI-----IVTQ-----GEVVIPEGV 823 |
| Db | 146 IPAGTDISTLNTNIEQIEILKGPFSIYGTNAMGGVNNIITHKSKDHGNSVLEFGS 205 |
| Qy | 824 YDYCITNPEPAGSKWMIADGGNQPARVDD--FTPEAGKKYFTWRRAGMGDGTMEVEDD 882 |
| Db | 206 Y-----QTNAGSFNLGG-----RPEDIFSPD-----LSLGLDKQNKD- 237 |
| Qy | 883 SPASYYTVTVRDG-----TKIKEGLTATTFEEDGVAAGNHEYCVE-----VKYTAG 928 |
| Db | 238 -----YKGTGNNFLSLSKLEEIVDYNATNKKMKGS-DYTVATGLRLFGIDFTPE 287 |
| Qy | 929 VSPKVKCDV-----TVEGSNEFPAPQNLTGSSVGQVKYTLKWDAPNGTTPNPNPNPNPG 982 |
| Db | 288 WSLNLYQNVLGDAIPVGGSENEFPAPQNLTGSSVGQVKYTLKWDAPNGTTPNPNPNPNPG 347 |
| Qy | 983 TTLSSEFENGIPASWKTIIDADGDGHGKWPAGNAPGIAGVNSNGCVYSSESGLGGTGLVLTDP 1042 |
| Db | 348 TTLSSEFENGIPASWKTIIDADGDGHGKWPAGNAPGIAGVNSNGCVYSSESGLGGTGLVLTDP 407 |
| Qy | 1043 NYLTLPALDLPNGGKLTFWCAQDANYASBEHYAYVYASSTGNDASNFNALLEETITAKGV 1102 |
| Db | 408 NYLTLPALDLPNGGKLTFWCAQDANYASBEHYAYVYASSTGNDASNFNALLEETITAKGV 467 |
| Qy | 1103 RSPKAIRGRIGQTRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVIKANGKRAADFTET 1162 |
| Db | 468 RSPKAIRGRIGQTRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVIKANGKRAADFTET 527 |
| Qy | 1163 FESESTGEAPAEWTTIDADGDGQGWCLSSGOLDWLTAHGGSNVVSFSWNGMALNDPNY 1222 |
| Db | 528 FESESTGEAPAEWTTIDADGDGQGWCLSSGOLDWLTAHGGSNVVSFSWNGMALNDPNY 587 |
| Qy | 1223 LISKDVGTGATKVYYIYAVNDGFPDGHYAVMISKTGTNAGDFTVVFETPENGKKGARFG 1282 |
| Db | 588 LISKDVGTGATKVYYIYAVNDGFPDGHYAVMISKTGTNAGDFTVVFETPENGKKGARFG 647 |
| Qy | 1283 LSTEANGAKPQSWIERTVLDLPAGTKYVAFRHVNCSDNLNILLDDIOFTWGGSPPTPDYDT 1342 |
| Db | 648 LSTEANGAKPQSWIERTVLDLPAGTKYVAFRHVNCSDNLNILLDDIOFTWGGSPPTPDYDT 707 |
| Qy | 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCEVKYTAGVSPKKCVYDVTVNQTQFNPVQ 1402 |
| Db | 708 YTVYRDGTKIKEGLTETTFEEDGVATGNHEYCEVKYTAGVSPKKCVNVTNQTQFNPVQ 767 |
| Qy | 1403 NLTAEQAPNSMDAILKWNAPASKRAEVLNEDFPENGIPASMKTIADADGDGNNWTTTPPPGG 1462 |
| Db | 768 NLTAEQAPNSMDAILKWNAPASKRAEVLNEDFPENGIPASMKTIADADGDGNNWTTTPPPGG 827 |
| Qy | 1463 SSFAGHNSATCVSSASHINPEGQPNPNLYVTPELSLPGGGTLTFWVCAQDANYASEHYA 1522 |
| Db | 828 SSFAGHNSATCVSSASHINPEGQPNPNLYVTPELSLPGGGTLTFWVCAQDANYASEHYA 887 |
| Qy | 1523 VYASSTGNDASNFANALLEEVLITAKTVTAPEAIRGRTRAGTWYQKTVQLPAGTKYVAFR 1582 |
| Db | 888 VYASSTGNDASNFANALLEEVLITAKTVTAPEAIRGRTRAGTWYQKTVQLPAGTKYVAFR 947 |
| Qy | 1583 HFGCTDFWNLDDVLTSGNAPSYYTIIYRNNTQIASGWTETTYRDPDLATGFYTYGVK 1642 |
| Db | 948 HFGCTDFWNLDDVLTSGNAPSYYTIIYRNNTQIASGWTETTYRDPDLATGFYTYGVK 1007 |

| | | | | |
|---|---|---|---|------|
| Qy | 1643 | VYVNGSSA | ETATLNITSLADVTQAQPYTLTVVVGKTIITVTCQGEAMIDMNGRRLAAGR | 1700 |
| Db | 1008 | VYVNGSSA | ETATLNITSLADVTQAQPYTLTVVVGKTIITVTCQGEAMIDMNGRRLAAGR | 1067 |
| Qy | 1703 | NTVYVTAQGHYAV | VMVVDGKSVEKLVAK | 1732 |
| Db | 1068 | NTVYVTAQGHYAV | VMVVDGKSVEKLVAK | 1097 |
| RESULT 12 | | | | |
| Q9R9B7 | | PRELIMINARY; | PRT; 1687 AA. | |
| ID | Q9R9B7 | AC | Q9R9B7; | |
| DT | 01-MAY-2000 | (T-EMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 | (T-EMBLrel. 13, Last sequence update) | | |
| DT | 01-OCT-2003 | (T-EMBLrel. 25, Last annotation update) | | |
| DE | Hemagglutinin/protease. | | | |
| GN | HAGE. | | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales; | | | |
| OC | Porphyromonadaceae; Porphyromonas. | | | |
| OX | NCBI_TaxID=837; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=381; | | | |
| RA | Han N., Dong H., Progulskie-Fox A.; | | | |
| RT | "Cloning and characterization of hAge from P. gingivalis 381."; | | | |
| EL | Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases. | | | |
| DR | EMBL; AF026946; AA001810.1; -. | | | |
| DR | HSSP; P95493; ICVR. | | | |
| DR | MEROPS; C25.001; -. | | | |
| DR | GO; GO:0005524; F:ATP binding; IEA. | | | |
| DR | GO; GO:0008234; F:cysteine-type peptidase activity; IEA. | | | |
| DR | GO; GO:0003910; F:DNA ligase (ATP) activity; IEA. | | | |
| DR | GO; GO:0006310; F:DNA recombination; IEA. | | | |
| DR | GO; GO:0006281; P:DNA repair; IEA. | | | |
| DR | GO; GO:0006260; P:DNA replication; IEA. | | | |
| DR | GO; GO:0006508; P:proteolysis and peptidolysis; IEA. | | | |
| DR | InterPro; IPR000977; DNA ligase. | | | |
| DR | InterPro; IPR007110; Ig-like. | | | |
| DR | InterPro; IPR001769; Peptidase C25. | | | |
| DR | InterPro; IPR005536; Peptidase C25 C. | | | |
| DR | Pfam; PF01364; Peptidase C25; 1. | | | |
| DR | Pfam; PF03785; Peptidase C25 C; 1. | | | |
| DR | PROSITE; PS00697; DNA_LIGASE_A1; 1. | | | |
| KN | Protease. | | | |
| SQ | SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64; | | | |
| Query Match 46.2%; Score 4245; DB 2; Length 1687; | | | | |
| Best Local Similarity 51.6%; Pred. No. 3 5e-211; | | | | |
| Matches 920; Conservative 197; Mismatches 443; Indels 224; Gaps 44; | | | | |
| Qy | 47 | FSFNEVELIKVETKGTGFASVSI | PGAPPTGCVGSPVPAVRKLIIVAGATPVVRVKS-- | 104 |
| Db | 30 | FRMDNLKFTVEVTPKGAQVPT | TEGVNLSKGMPTLPILSRSLAVSDTRMKVEVVS | 89 |
| Qy | 105 | FTRQVYSLNQYSEKLMHPQSP | SKSDDEKPYFVYNAAYARKGVGOELTQVEMLGT | 164 |
| Db | 90 | FIEKKNL----- | IAPSKGMIMRNEPKKIPYV-GKSYQNKFFPG | 141 |
| Qy | 165 | RGVRIALITNPVQYDVAWNLK | VRNNIEI---EVSFGQADEVATQRLYDASFPYFETA | 221 |
| Db | 142 | RDVRGQVWFAPLOYNPVTKL | RIYETIIVAVSETSEQKNIINKGTAG----- | 196 |
| Qy | 222 | YKOLFNRDVTYDHGDIYNTPV | -----RMLVVAGAKFEALKPWLTWKAKGFYLDVHYT | 275 |
| Db | 197 | YKRMF----- | MNYEPGRYTPVEBKQNGRMIVIVAKKYEGDKDFVDW | 251 |
| Qy | 276 | DEAEVGTWNASIKAFTHKKY | ----NDGLAASAPFLAIVGDTDIVSGE-- | 329 |
| Db | 252 | EDIASPVTAALQQFVKQYKE | GND-----LTVLLVGVGDKDIPAKITPGIKSDQV- | 303 |
| Qy | 330 | DLYSYADVGDYFPEMYTFRMS | ASSPEELNIIDKVLNWEKATWPKDSYLEKVLIIAGADY | 389 |

Db 304 --YGVGNNDHYNEVIFGRFSCSKEDLKTQIDRTIHYERNITTEDKWLQALCIASAE 361
QY 390 SWNSQVGOPTIKY-CMOYVYVNEQHCYTDVYNVLPKPYTGCY-----SHLNTGVSF 438
Db 362 GPSADNGSDIOHENVIANLITQYGTIKI-----CYDPGVTPKNIIDAFNGGISL 413
QY 439 ANYTAHSETAWAPLLTTSOLKALTNKDKYFLAIGNCCITAQPDYVOPCFGEVITRV-- 496
Db 414 VNYTGHSETANGTSHFGTTHVKQLTNSQLPFFIDVACVNGDFLPSMPCFAELMRAQK 473
QY 497 --KKGAYAYIGSSPNSWGBDYWVGANAVFGVQPTFECTSMGSDYDTELEDSYNFVN 554
Db 474 DGKPTGTVAIIASTINQSW-----ASPMRG-----QDEMNEI- 505
QY 555 SIMWAGNLAATHAGNI-----GNITHIGHYWEAYH-----VLGDSWMPYRAM 599
Db 506 -----LCEHPNIIKRTFGVTWNGMFMVKEYKKDGERKMLDTWTVFGDPSSLVRLV 558
QY 600 PKNITTYLTPASLPONQASYSIQAS-AGSYVAISKDGLVLYGTGVANASGVATVSMTKOITE 658
Db 559 PTKMQVTAPAQINLTDSVNSVDYNGAIAIISANGRMFGSAVVE-NGTAIINIT-GLTN 616
QY 659 NGYDVVITRNYLPVIKIOV-GEPSYQVPSNLTATTOQKVTYLKWEAPSAK---KAE 714
Db 617 ESTLTLTVVGYNKETVKTINTNGEPNPQVPSNLTATTOQKVTYLKWDADSTKTNTN 676
QY 715 GSREVKEIGDLFTVIBPAND-VRANAKVVLADNVDGNTGYQFLDADHNTFGSVIP 773
Db 677 TARSVDGIRELVLSVSDAPELLASGQAEIVLEADHVDNNGSGYQIILLDADHDYGOVIP 736
QY 774 A-TGRLFTG-TASSNLYSANFEXLVPANADPVVTTQNIITVGOGEVIVPGVYDICTNP 831
Db 737 SDTHLWPNCSVPANLF-APPEYVTPENADPSCSPTNIMDGTASVNPAGTYDFAAAP 795
QY 832 EPASGKMKIAGDGNQARVDYDFEAGKKYFTTMRAGMGDGTMDVEDDSPASYTYV 891
Db 796 Q-ANAKIWIAGQ---PTKEDDYFEAGKKYHFLMKWGSQDGTETATISEGGSDYTYV 851
QY 892 YRDGTKIKEGLTATTFEEDGVAAGNHEVCVVKYTAGVSPKVKDVTVEGSENEAPVQNL 951
Db 852 YRDGTKIKEGLTATTFEEDGVAAGNHEVCVVKYTAGVSPKVKDVTVEGSENEAPVQNL 911
QY 952 TGSVGOQVTLKWDAPNGTNPNNPNPNPFG---TTLSEFENGIPASWKTIADADGQGH 1008
Db 912 TGSVGOQVTLKWDAPNGTNPNNPNPNPFGTTLSEFENGIPASWKTIADADGQGH 971
QY 1009 WKPCNAPGIAGYNSGCVYSESFGIGGIGVLTDPNYLITPALDLPNGGKLTFFWVCAQDAN 1068
Db 972 WKPCNAPGIAGYNSGCVYSESFGIGGIGVLTDPNYLITPALDLPNGGKLTFFWVCAQDAN 1031
QY 1069 YASEHYAVYASSTGNDASNTFALLEETITAKGVRSKAIKRGITQGTWRKOTVDLPAGTK 1128
Db 1032 YASEHYAVYASSTGNDASNTFALLEETITAKGVRSKAIKRGITQGTWRKOTVDLPAGTK 1091
QY 1129 YVAFRHQSTDMFVIDLDEVEIKANGKADFTETFEESTHGEAPAEWTTTIDADGGQWL 1188
Db 1092 YVAFRHQSTDMFVIDLDEVEIKANGKADFTETFEESTHGEAPAEWTTTIDADGGQWL 1151
QY 1189 CLSSGQDLWLTAGGNSVNSFSGNGALPNPNYLSKDVTKATKYYAVYAVNDGFGPDH 1248
Db 1152 CLSSGQDLWLTAGGNSVNSFSGNGALPNPNYLSKDVTKATKYYAVYAVNDGFGPDH 1211
QY 1249 YAVMISKTGTNAGDFTVVFETPNNGKNGARFGLSTEANGAKPQSVMIERTVDLPAGTK 1308
Db 1212 YAVMISKTGTNAGDFTVVFETPNNGKNGARFGLSTEANGAKPQSVMIERTVDLPAGTK 1271
QY 1309 YVAFRHVNSCDLNYILLDDIQTWGGSPPTDITYTYVYRDGTKIKEGLTETTFEEDGVAT 1368
Db 1272 YVAFRHVNSCDLNYILLDDIQTWGGSPPTDITYTYVYRDGTKIKEGLTETTFEEDGVAT 1331
QY 1369 GNHEYCVVEKYTAGVSPKVKDVTVNSTQFNPNQNLTAEOAPNSMAILKWNAPASKRAE 1428

Db 1332 GNHEYCVVEKYTAGVSPKVKDVTVNSTQFNPNQNLKQAQ--PDGDDVVLKWEAPSAKTE 1389
QY 1429 -----VLNEDFENGIPASWKTIADADGNNWTTTPP 1459
Db 1390 GSREVKEIGDLFTVIBPANDVRANEAKVVLADNVDGNTGYQFL-LDADHNTFGSVIP 1448
QY 1460 PGSSSFAGHNSAICVSSASHINPEG--PQMPD-----NYLVT--PELSLPGGTLTFWV 1509
Db 1449 ATGFLFTGTASSNLYSA-----NFEYLIPANADPVVTTQNIITVGOGEVIVPGG---YDY 1501
QY 1510 CAODANYASHYAVYASSTGNDASNTFALLEEVLTAKTVVTAPETARGTRAQGTWQKT 1569
Db 1502 CITNPEFASGKMWI-AGDGNQPARYDDTFE----- 1532
QY 1570 VQLPAGTKYV-AFRHFCCTDFFWINLDDVVTGNSNAPSYTYTYRNNTQIASGVTTETTYR 1628
Db 1533 ---AGKKYFTTMRAGMGDGT---TDMVEDDSPASYTYTYVYRDGTKIKEGLTETTYR 1583
QY 1629 DPDLATGFTYGVKVVYPNGESALETATLNTSLADVTACKPYTLTVVGKTIITVTCGEA 1688
Db 1584 DAGMSAQSHEYCEVKTAAAGVSPKVCVDYIPDGVADVTAQPKYTLTVVGKTIITVTCGEA 1643
QY 1689 MIXDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK 1732
Db 1644 MIXDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK 1687
RESULT 13
P96966 PRELIMINARY; PRT; 364 AA.
AC P96966;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Cysteine protease.
GN KGP-381.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA Han N., Lepine G., Whitlock J., Wojciechowski L., Progulake-Fox A.;
RT "Cloning, sequencing and characterization of hagd, a member of the
RT Harep multigene family in Porphyromonas gingivalis.";
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U69469; AAB49690.1; --
DR MEROPS; C25.002; --
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001769; Peptidase_C25.
DR Pfam; PF01364; Peptidase_C25; I.
KW Protease.
SQ SEQUENCE 364 AA; 40161 MW; 33DB6A81CF57557B CRC64;
Query Match 20.2%; Score 1854; DB 2; Length 364;
Best Local Similarity 100.0%; Pred. No. 3e-88; 0; Indels 0; Gaps 0;
Matches 363; Conservative 0; Mismatches 0;
QY 1 MRKLLLLIAASLLGLVLAQSAKIKLDAPTTCTTNNSTFKQDFASFSNEVELTKVETK 60
Db 1 MRKLLLLIAASLLGLVLAQSAKIKLDAPTTCTTNNSTFKQDFASFSNEVELTKVETK 60
QY 61 GGTFAVSITPGAPPTGEVGSPEVPAVRKLIAPVVGATPVVRKSFTEQVYSLNQYSEKL 120
Db 61 GGTFAVSITPGAPPTGEVGSPEVPAVRKLIAPVVGATPVVRKSFTEQVYSLNQYSEKL 120
QY 121 MPHOPSMKSDDEPKVPFVYVYNAARVAKGVQGLTQVEMLTGMGRVTAALTINPVQYD 180
Db 121 MPHOPSMKSDDEPKVPFVYVYNAARVAKGVQGLTQVEMLTGMGRVTAALTINPVQYD 180
QY 181 VVANQLKVRNNIEIEVSFQGADEVATORLYDASFSFPYFETAYKQLFNRDVYTHDGLYNT 240

```

Db 181 VVANQLKVRNNIEIEVSFGQAEVATQRLYDASFSPYETAYKQLFNRDVTYDHGDLNT 240
QY 241 PVRMLVAGAKEKALKPMLTWKAQGFYLDVHYTDEAVGTTNASIKAIFIHKYNDGLA 300
Db 241 PVRMLVAGAKEKALKPMLTWKAQGFYLDVHYTDEAVGTTNASIKAIFIHKYNDGLA 300
QY 301 ASAAVFALVGDVTDVISGEKKTCKVTDLVYSVADGDFPFEMTFRMSASSPELTNI 360
Db 301 ASAAVFALVGDVTDVISGEKKTCKVTDLVYSVADGDFPFEMTFRMSASSPELTNI 360
QY 361 IDK 363
Db 361 IDK 363

RESULT 14
Q9F4J0 PRELIMINARY; PRT; 925 AA.
AC Q9F4J0;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative outer membrane protein PG57.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT "Identification of vaccine candidates from genomic analysis of
RT Porphyromonas gingivalis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007522; AAG24228.1; --
DR InterPro; IPR003961; FN_III.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 925 AA; 103632 MW; 5FF2198D6914DAE8 CRC64;

```

```

Query Match 10.7%; Score 980.5; DB 2; Length 925;
Best Local Similarity 28.4%; Pred. No. 2.4e-42;
Matches 289; Conservative 109; Mismatches 264; Indels 357; Gaps 42;

QY 960 VTLKWA-PNGFNENPNPN------PGTLESFENG-IPASWKTIDADGGHGMK 1010
Db 16 ILLSWAATNPTAQISGNASCLAPQPDILIESFENGVPNGWLEIDADAGATWG 75
QY 1011 PGNAPGIAGYNSG-CVYSEFGLGIGVLTDPNLYITPALDPNGGKLTFWVCQADANY 1069
Db 76 SPGSFSPVPGHGLCTYSHRS-----GISTAGNVLITP--NIEGAKRVKVCNQ---Y 126
QY 1070 AS--PHYAVASSTGNDASNFNALLEITAKGVRSKAIKRGITQGTWRQKTVLDLPGT 1127
Db 127 STNPHYAVMVSTGTATIEDFY-LLFDDSIK-PTPL-----VWRRRIVDLPEGT 175
QY 1128 KYVAFRHFGSTDMF--YIDLDEV----- 1148
Db 1149 -----EIKANGK-----RADFTETFE 1164
QY 236 EKGNGEELQLSGYNIYANGTLLAQIKDVLSLEYVDSTYSLRDNPLQVEYCVTAVDIESIE 295
Db 1165 SST-----HGEAPAEWTTIDAGDGGWCLSSGQ-----LDWLTAHG 1202

```

```

Db 296 SSTVCGTLHYATDAILYENFENGVPNGWLVTDADGGSW-----GHVLYADAPPEN 350
QY 1203 GSNVVSFSW-NGMA-LNPDNYLSKDVGTGATKVYYAVND-GPFGDHYAVMISKTGTN 1259
Db 351 GGHCSLSASVYVPGIVTDPNLYITPKVEGAKRVKYSTQDANWAAEHYAVMASTTGA 410
QY 1260 AGDFTVVEEPENGINKGAPGLSTEANGAKPQSWIERTVDLPAGIKYVAFRINCSL 1319
Db 411 VGDFFVILFEETNT-----AKPTGAWYERTINLPEGTKYIAWRHYNCTD 453
QY 1320 LNYILLDDIQF--TMGGSPTP--TDYT----- 1342
Db 454 IYFLKLDLITVFGTAPASEPEPTDFVSLIENNKGLKWNYPNGYEPDKTDDKDLQLAG 513
QY 1343 YTVYRDGT-----KKEGLTET--TFEEDGATGNHEYCVVEYKTAGVSPKCVDTVT 1393
Db 514 YNIYANGSLLVHIQDPTVLEIDETYSRDDQVEV---EYCVTAVYNDNIESQSVCCKLI 570
QY 1394 NSTQFNPVQNLTAEOAPNSMDAILKWNAPASKRAEVLNEDFENG-IPASWKTIDADGDN 1452
Db 571 YDSQSD-----IILYEGFEAGSIPEGNLLIDADGDNV 602
QY 1453 NWTTPPGSSSFAGHNSAICVSSASHINFEQPNDNYLVTPELSLPGGGTITFWVCAQ 1512
Db 603 NWDYYP---WTMYGHDSEKCIASPSYLMIGVLTDPNLYITPRL--GAKLVKVVWSAQ 656
QY 1513 DANYASEHYAVYASSTGNDASNFANALLEVLTAKTVVTAPPAIRGTRAGTQWYQKTVOL 1572
Db 657 DAVYSAEHYAVMVSTGTAVEDEV-LLFEETMTAK-----ANGAWYERTITL 702
QY 1573 PAGTKYVAFRHFGCTDFWNLDDVVITSGNAPSYYTIYRNNTQIASGVTE----- 1624
Db 703 PAGTKYIAWRHYDCHDMFILLDDI-----TVYRSTETVPPEPTDFVVSLEN 750
QY 1625 -----TYRDPDLATGTYTVKVVYPNGESAVESATETATNITSLADVT- 1666
Db 751 NKGRKWNYPNGYEPDKTDDKPLQLTGVN-----IYANGSLLVHIQDPTVLEIDET 804
QY 1667 -----AQKPYLTIV-----GKTIIV--TCQGEAM-----YDMNGR 1696
Db 805 SSRDQGVEMECVTAVYNDNIESQSVCCKLYITITSLDNIQSDTSLKTIYPNPASVVR 864
QY 1697 RLAAGRNTV-VYTAQG-----GHYAVMVVDGKSVKELAVK 1732
Db 865 GLRSKSTIELYNALGICILREETHSEKTEIDVSLNDGVILKVVGGNKTTEKVEIK 923

RESULT 15
Q9KIB3 PRELIMINARY; PRT; 312 AA.
AC Q9KIB3;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

```

DR EMEL; AF237555; AAF81413.1; -
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;
Query Match 7.9%; Score 722; DB 2; Length 312;
Best Local Similarity 47.5%; Pred. No. 1.2e-29;
Matches 150; Conservative 49; Mismatches 97; Indels 20; Gaps 7;
QY 1430 LNEDEFNGIPASWKTIADGDGNNWTTTPPGSSSFAGHNSAICVSSASHINPEGPONPD 1489
Db 4 LSESFESGIFAIWKTIADGDGNNWHL-----TNFTG-QSGLCVSSASIIGGVGALTPD 57
QY 1490 NYLVTPELSLPGGG--TLTFWVCAQDANYASEHYAVYASSTGNDASNANFALLEEVLTA 1547
Db 58 NYLITPELKLPTDALVEIIVVCTQDLTAPSEHYAVYSSSTGNNADFNLLYEETLTAK 117
QY 1548 TVVTAPAEIRGTRAQGTWYQKTVQLPAGTKVAFRHEGCTDFPWINLDDVVITSGNAP-- 1605
Db 118 R-IQSPELIRGNRTQGVYQKVVLPNDTKYVAFRHFNSTDNFNLDELVSILYTPLPRR 176
QY 1606 -----SYTYIYRNNTQIASGVTTETTRDPDLATGFYTYGVKVYPNGESALETATLN 1658
Db 177 APCPHPGGYTYSVFRDQOKIASGLSALAYTDTVPYGTQDYCVQVNYLQGSYKVKCNIV 236
QY 1659 ITSADV-TAQKPYTLTWVGKTIITWC-QGEAMTYDMNGRRLAAGRNTVVYTAQGGHYAV 1716
Db 237 VANSANIYGADKPPALTWGKTIIVASAFKGEITLYDIRGLIASGCDTLRYKKAENGFYLI 296
QY 1717 MVVVDGKSYYVEKLAVK 1732
Db 297 KIQVNGTVTYTEKIQIQ 312

Search completed: May 18, 2004, 11:35:11
Job time : 75.1557 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:30:39 ; Search time 28.7501 Seconds
(without alignments)
3110.116 Million cell updates/sec

Title: US-08-353-485-10

Perfect score: 9179

Sequence: 1 MKKLLILIAASLLGVGLVLAQ.....HYAVMVVDGKSVVEKLVK 1732

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:**

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:**

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:**

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:**

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:**

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 9179 | 100.0 | 1732 | 2 | US-08-570-311-10 |
| 2 | 9179 | 100.0 | 1732 | 2 | US-08-353-485-10 |
| 3 | 9167 | 99.9 | 1732 | 4 | US-09-066-330-11 |
| 4 | 6151.5 | 67.0 | 1358 | 2 | US-08-570-311-27 |
| 5 | 4811 | 52.4 | 942 | 1 | US-08-141-324-14 |
| 6 | 4811 | 52.4 | 942 | 1 | US-08-541-902-14 |
| 7 | 4673.5 | 50.9 | 1087 | 2 | US-08-570-311-8 |
| 8 | 4673.5 | 50.9 | 1087 | 2 | US-08-353-485-8 |
| 9 | 4272 | 46.5 | 1706 | 4 | US-09-066-330-10 |
| 10 | 4270 | 46.5 | 1704 | 3 | US-08-336-308A-10 |
| 11 | 4270 | 46.5 | 1704 | 3 | US-08-822-324-6 |
| 12 | 4270 | 46.5 | 1704 | 3 | US-09-490-931-10 |
| 13 | 4245 | 46.2 | 1687 | 2 | US-08-570-311-29 |
| 14 | 4217 | 45.9 | 2628 | 2 | US-08-570-311-14 |
| 15 | 4135.5 | 45.1 | 1477 | 4 | US-09-482-500A-1 |
| 16 | 2565 | 27.9 | 509 | 3 | US-08-822-324-8 |
| 17 | 2034.5 | 22.2 | 456 | 2 | US-08-570-311-18 |
| 18 | 2034.5 | 22.2 | 456 | 2 | US-08-570-311-20 |
| 19 | 2025.5 | 22.1 | 439 | 2 | US-08-570-311-22 |
| 20 | 2011.5 | 21.9 | 450 | 2 | US-08-570-311-16 |
| 21 | 1854 | 20.2 | 364 | 2 | US-08-570-311-26 |
| 22 | 812.5 | 8.9 | 497 | 2 | US-08-570-311-2 |
| 23 | 812.5 | 8.9 | 497 | 2 | US-08-353-485-2 |
| 24 | 698 | 7.6 | 737 | 1 | US-08-119-361-5 |
| 25 | 698 | 7.6 | 737 | 3 | US-08-336-308A-4 |
| 26 | 698 | 7.6 | 737 | 3 | US-08-822-324-4 |
| 27 | 698 | 7.6 | 737 | 3 | US-09-490-931-4 |

| | | | | | | |
|----|-------|-----|------|---|----------------------|-------------------|
| 28 | 526 | 5.7 | 492 | 4 | US-09-482-500A-2 | Sequence 2, Appli |
| 29 | 412.5 | 4.5 | 507 | 4 | US-09-482-500A-3 | Sequence 3, Appli |
| 30 | 309 | 3.4 | 1747 | 4 | US-09-134-000C-5999 | Sequence 5999, Ap |
| 31 | 306 | 3.3 | 3892 | 4 | US-09-328-352-5503 | Sequence 5503, Ap |
| 32 | 304.5 | 3.3 | 2736 | 4 | US-09-252-991A-30227 | Sequence 30227, A |
| 33 | 293 | 3.2 | 1638 | 4 | US-09-071-035-258 | Sequence 258, App |
| 34 | 293 | 3.2 | 1638 | 4 | US-09-071-035-262 | Sequence 262, App |
| 35 | 293 | 3.2 | 1638 | 4 | US-09-071-035-266 | Sequence 266, App |
| 36 | 277 | 3.0 | 2315 | 4 | US-09-543-681A-5434 | Sequence 5434, Ap |
| 37 | 250.5 | 2.7 | 2411 | 4 | US-09-268-347-36 | Sequence 36, Appl |
| 38 | 246.5 | 2.7 | 3290 | 4 | US-09-328-352-5486 | Sequence 5486, Ap |
| 39 | 237.5 | 2.6 | 2353 | 3 | US-09-377-155-33 | Sequence 33, Appl |
| 40 | 237.5 | 2.6 | 2353 | 3 | US-08-913-942-4 | Sequence 4, Appli |
| 41 | 237.5 | 2.6 | 2353 | 4 | US-09-669-974-33 | Sequence 33, Appl |
| 42 | 237.5 | 2.6 | 2353 | 4 | US-09-797-862-33 | Sequence 33, Appl |
| 43 | 236 | 2.6 | 49 | 3 | US-08-822-324-18 | Sequence 18, Appl |
| 44 | 236 | 2.6 | 952 | 4 | US-09-107-532A-4706 | Sequence 4706, Ap |
| 45 | 235.5 | 2.6 | 1912 | 1 | US-08-409-995-4 | Sequence 4, Appli |

ALIGNMENTS

RESULT 1
US-08-570-311-10
; Sequence 10, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 1732 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-570-311-10

Query Match      100.0%; Score 9179; DB 2; Length 1732;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLLLLAASLLGVLGAQSAKIKLDAPTRTTCTNNNSFKQDFASFSFNEVELTKVETK 60
Db 1 MRKLLLLAASLLGVLGAQSAKIKLDAPTRTTCTNNNSFKQDFASFSFNEVELTKVETK 60

QY 61 GGTASVSIPGAFPTGEVSPVPAVKLIAPVPGATPVVRKSTTEOVYSINQVSEKL 120
Db 61 GGTASVSIPGAFPTGEVSPVPAVKLIAPVPGATPVVRKSTTEOVYSINQVSEKL 120

QY 121 MPHQSMSKSDDEPKVPFVYNAAYARKGFVQELTQVEMLTMRGVRIAAALTINPVQYD 180
Db 121 MPHQSMSKSDDEPKVPFVYNAAYARKGFVQELTQVEMLTMRGVRIAAALTINPVQYD 180

QY 181 VVANQLKVRNNLEIVSFQGADEVATQRLYDASFSFYFETAYKQLENRDVYTDHGDLINT 240
Db 181 VVANQLKVRNNLEIVSFQGADEVATQRLYDASFSFYFETAYKQLENRDVYTDHGDLINT 240

QY 241 FVRMLVAGAKFKEALKPWLTKWAKQGFYLDVHYTDEAEVGTINASIKAFTHKKYNDGLA 300
Db 241 FVRMLVAGAKFKEALKPWLTKWAKQGFYLDVHYTDEAEVGTINASIKAFTHKKYNDGLA 300

QY 301 ASAPVFALVGDGTDVISEGKKTCKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360
Db 301 ASAPVFALVGDGTDVISEGKKTCKVTDLYYSAVDGDYFPEMYTFRMSASSPEELTNI 360

QY 361 IDKVLMEKATMPDKSYLEKVLITAGADYSNNSQVQPTIKYQMYYNQHGHTDVYNY 420
Db 361 IDKVLMEKATMPDKSYLEKVLITAGADYSNNSQVQPTIKYQMYYNQHGHTDVYNY 420

QY 421 LKAPYTCYSHLNTGVSFANYTAHGSSETAWADPLLTTSOLKALTNKDYFLAIGNCCITA 480
Db 421 LKAPYTCYSHLNTGVSFANYTAHGSSETAWADPLLTTSOLKALTNKDYFLAIGNCCITA 480

QY 481 QFDVYQPCFGEVITRVKEKGAAYIVIGSSPNSYWGEDYYWSVGANAFVGVQPTFEGTSMGS 540
Db 481 QFDVYQPCFGEVITRVKEKGAAYIVIGSSPNSYWGEDYYWSVGANAFVGVQPTFEGTSMGS 540

QY 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNICNITHIGAHYTWAEYHVLGDGVSMPYRAMP 600
Db 541 YDATFLEDSYNTVNSIMWAGNLAATHAGNICNITHIGAHYTWAEYHVLGDGVSMPYRAMP 600

QY 601 KTNNTYTLPASLPQOASYSIQASAGSYVAISKDGLVYGTGVANASGVATVSMTKQITENG 660
Db 601 KTNNTYTLPASLPQOASYSIQASAGSYVAISKDGLVYGTGVANASGVATVSMTKQITENG 660

QY 661 NYDVVITRESNVLPIVKIOVCEPSYQVSNLTATTOGQKVTLLKWEAPSAKKAEGSREVK 720
Db 661 NYDVVITRESNVLPIVKIOVCEPSYQVSNLTATTOGQKVTLLKWEAPSAKKAEGSREVK 720

QY 721 RIGDGLFVTIETPANDVRANEAKVLAADNVWMDNTGYQFLDADHNTFGSVIPATGPLFT 780
Db 721 RIGDGLFVTIETPANDVRANEAKVLAADNVWMDNTGYQFLDADHNTFGSVIPATGPLFT 780

QY 781 GTASSNLYSANFEYLVPANAPVVTQNIIVTGQGEVVIIPGVYDYCITNPEPASGRKWI 840
Db 781 GTASSNLYSANFEYLVPANAPVVTQNIIVTGQGEVVIIPGVYDYCITNPEPASGRKWI 840

QY 841 AGDGNOPARYDDFTFAGKKYITMRAGNGDGTMEVEDDSDSPASYTYTVYRDGTIKI 900
Db 841 AGDGNOPARYDDFTFAGKKYITMRAGNGDGTMEVEDDSDSPASYTYTVYRDGTIKI 900

QY 901 GLTATTTEEDGVAAGNHEYCVVEKYTAGVSPKVCCKVTVEGSNEFAPVQNLTGSSVQKV 960
Db 901 GLTATTTEEDGVAAGNHEYCVVEKYTAGVSPKVCCKVTVEGSNEFAPVQNLTGSSVQKV 960

QY 961 TLKWDAPNGTNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNP 1020
Db 961 TLKWDAPNGTNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNP 1020

QY 1021 NSNGCVTSSEFGLGGIGVLTDPDNYLITPALDPNGGKLTFWCAQDANYASEHVAVYASS 1080
Db 1021 NSNGCVTSSEFGLGGIGVLTDPDNYLITPALDPNGGKLTFWCAQDANYASEHVAVYASS 1080

QY 1081 TGNDASNTNALLBETITAKGVRSPKAIIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDTM 1140
Db 1081 TGNDASNTNALLBETITAKGVRSPKAIIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSDTM 1140

QY 1141 FYIDLDEVEIKANGKRADFTETTESSTHGEAPAEWTTIDADGGQGGWMLCLSSGOLDMLTA 1200
Db 1141 FYIDLDEVEIKANGKRADFTETTESSTHGEAPAEWTTIDADGGQGGWMLCLSSGOLDMLTA 1200

QY 1201 HGGSNVYSSFSWNGMALNPONYLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGNA 1260
Db 1201 HGGSNVYSSFSWNGMALNPONYLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGNA 1260

QY 1261 GDFTVVFEETPENGINKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDL 1320
Db 1261 GDFTVVFEETPENGINKGGARFGLSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDL 1320

QY 1321 NYILLDDIQTMTGSGPTDYTYTVYRDGTKIKEGLTETTFEEDGVAIGHHEYCVVEKYT 1380
Db 1321 NYILLDDIQTMTGSGPTDYTYTVYRDGTKIKEGLTETTFEEDGVAIGHHEYCVVEKYT 1380

QY 1381 AGVSPKCKVDVTNOSTQFNQVNLTAQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440
Db 1381 AGVSPKCKVDVTNOSTQFNQVNLTAQAPNSMDAILKWNAPASKRAEVLNEDFENGIPA 1440

QY 1441 SWKTIADGGGNKNTTTPPGGSGSFAGHNSAICVSSASHINFEQPNPDNYLVTPELSLP 1500
Db 1441 SWKTIADGGGNKNTTTPPGGSGSFAGHNSAICVSSASHINFEQPNPDNYLVTPELSLP 1500

QY 1501 GGGTLTFWCAQDANYASEHVAVYASSSTGNDASNFANALLEEVLTAKTAVTAPEAIRGTR 1560
Db 1501 GGGTLTFWCAQDANYASEHVAVYASSSTGNDASNFANALLEEVLTAKTAVTAPEAIRGTR 1560

QY 1561 AQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLDDVVITSGNAPSXYTYIYRNNTOIAS 1620
Db 1561 AQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLDDVVITSGNAPSXYTYIYRNNTOIAS 1620

QY 1621 GVTETTYRDPDLATGFTYTGKVVYPNGESAIEATLNTSLADVTAQKPYTLTVVGKTI 1680
Db 1621 GVTETTYRDPDLATGFTYTGKVVYPNGESAIEATLNTSLADVTAQKPYTLTVVGKTI 1680

QY 1681 TVTTCQGEAMLYDNMGRRLAAGRNTVVYTAQGHYAVWVVDGKSYVEKLAVK 1732
Db 1681 TVTTCQGEAMLYDNMGRRLAAGRNTVVYTAQGHYAVWVVDGKSYVEKLAVK 1732

RESULT 2
US-08-353-485-10
; Sequence 10, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
; APPLICANT: Proguale-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whillock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
```

COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/353,485
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/647,119
 FILING DATE: 25-JAN-1991
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/241,640
 FILING DATE: 08-SEP-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Whitlock, Ted W.
 REGISTRATION NUMBER: 36,965
 REFERENCE/DOCKET NUMBER: UF15-C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (904) 375-8100
 TELEFAX: (904) 372-5800
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1732 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 us-08-353-485-10

Query Match 100.0%; Score 9179; DB 2; Length 1732;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRKLLLLIAASLLGVLYAQSAKIKLDAPTRTRTCTNNSFKQFDASFNEVELTKVETK 60
 DB 1 MRKLLLLIAASLLGVLYAQSAKIKLDAPTRTRTCTNNSFKQFDASFNEVELTKVETK 60
 QY 61 GGFASVISIPGAPFTGEGVSPVAVRKLTAIPVGATPVVRKVSFTQVYSLNQYSEKL 120
 DB 61 GGFASVISIPGAPFTGEGVSPVAVRKLTAIPVGATPVVRKVSFTQVYSLNQYSEKL 120
 QY 121 MPHQPMSKSDDEKVPFVYNAAYARKGFVQELTQVEMLTMRGVRIIAALTINPVQYD 180
 DB 121 MPHQPMSKSDDEKVPFVYNAAYARKGFVQELTQVEMLTMRGVRIIAALTINPVQYD 180
 QY 181 VVANQLKVRNNIEIVSFQGADEVATORLYDASFSYPFETAYKQLFNRDVTYDHGDLNT 240
 DB 181 VVANQLKVRNNIEIVSFQGADEVATORLYDASFSYPFETAYKQLFNRDVTYDHGDLNT 240
 QY 241 PVRLMVAGAKFKEALKPWLTKAOKGFYLDVHYTDEAEVGTNNASIKAFIHKKYNDGLA 300
 DB 241 PVRLMVAGAKFKEALKPWLTKAOKGFYLDVHYTDEAEVGTNNASIKAFIHKKYNDGLA 300
 QY 301 ASAPVFLALVGTDTVISEGKGGKTKKVTDLXYSAVDGDFPEMYTPRMSASSPEELTNI 360
 DB 301 ASAPVFLALVGTDTVISEGKGGKTKKVTDLXYSAVDGDFPEMYTPRMSASSPEELTNI 360
 QY 361 IDKVLMEKATMPDKSKYLEKVLIIAGADYSWNSQVQPTIKYGMQYNNQEHGVTDVYNY 420
 DB 361 IDKVLMEKATMPDKSKYLEKVLIIAGADYSWNSQVQPTIKYGMQYNNQEHGVTDVYNY 420
 QY 421 LKAPYTCYSHLNTGVSFANYTAGHSTAWADPLLITTSQKALTNKDKYFLAIGNCCITA 480
 DB 421 LKAPYTCYSHLNTGVSFANYTAGHSTAWADPLLITTSQKALTNKDKYFLAIGNCCITA 480
 QY 481 QFDYVQPCFGEVITRVKEKAYAYIGSSPNSWGEDYYSWGVANAVFGVQPTFEGTSMGS 540
 DB 481 QFDYVQPCFGEVITRVKEKAYAYIGSSPNSWGEDYYSWGVANAVFGVQPTFEGTSMGS 540

541 YDATEFLDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGVSMPYRAMP 600
 DB YDATEFLDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGVSMPYRAMP 600
 QY 601 KTNVYTLPASLPQNOASYSIOASAGSYVAISKDGLYGTGVANASGVATVSMTKQITENG 660
 DB 601 KTNVYTLPASLPQNOASYSIOASAGSYVAISKDGLYGTGVANASGVATVSMTKQITENG 660
 QY 661 NYDVIITRSNYLPIVKIQIIVGEPSPQVSNLTATTOGQKVTLKWEAPSAKKAEGSRV 720
 DB 661 NYDVIITRSNYLPIVKIQIIVGEPSPQVSNLTATTOGQKVTLKWEAPSAKKAEGSRV 720
 QY 721 RIGDGLFVTIIEPANDVRANEAKVVLAAADNVGDNNTGYOFLLDADHNTFGSVIPATGPLET 780
 DB 721 RIGDGLFVTIIEPANDVRANEAKVVLAAADNVGDNNTGYOFLLDADHNTFGSVIPATGPLET 780
 QY 781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTQGEVVIIPGVVYDICTNPEPASGKOWI 840
 DB 781 GTASSNLYSANFEYLVPANADPVVTTQNIIVTQGEVVIIPGVVYDICTNPEPASGKOWI 840
 QY 841 AGDGNQPARYDDFTFEAGKXYTFTMRACMGDGTMEVEDDSPASYTYTVYRDGTIKE 900
 DB 841 AGDGNQPARYDDFTFEAGKXYTFTMRACMGDGTMEVEDDSPASYTYTVYRDGTIKE 900
 QY 901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVKDVTVEGSNEFAPVONLTGSSVGQKV 960
 DB 901 GLTATTFEEDGVAAGNHEYCVEVKYTAGVSPKVKDVTVEGSNEFAPVONLTGSSVGQKV 960
 QY 961 TLKWDAPNGTNPNNPNNPNTLSESPFENGIPASWKTIIDADGDGHWKPGNAPGIAGY 1020
 DB 961 TLKWDAPNGTNPNNPNNPNTLSESPFENGIPASWKTIIDADGDGHWKPGNAPGIAGY 1020
 QY 1021 NSNGCVYSESGGLGGVLTDPDNYLITPALDLPNGKLTFWCAQDANYASEHYAVYASS 1080
 DB 1021 NSNGCVYSESGGLGGVLTDPDNYLITPALDLPNGKLTFWCAQDANYASEHYAVYASS 1080
 QY 1081 TGNDSNFTNALLBEITITAKGVSPKAIKRIIOGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
 DB 1081 TGNDSNFTNALLBEITITAKGVSPKAIKRIIOGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
 QY 1141 FYIDLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGDGQWCLCLSSGQDMLTA 1200
 DB 1141 FYIDLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGDGQWCLCLSSGQDMLTA 1200
 QY 1201 HGSNNVVSFSWNGMALNPNDYLI SKDVTGATKVKYVAVNDGFGPDGHYAVMI SKTGNA 1260
 DB 1201 HGSNNVVSFSWNGMALNPNDYLI SKDVTGATKVKYVAVNDGFGPDGHYAVMI SKTGNA 1260
 QY 1261 GDFTVVFEETPNKGGARFGLSTEANGAKPOSVMERTVDLPAGTKYVAFRHYNCSDL 1320
 DB 1261 GDFTVVFEETPNKGGARFGLSTEANGAKPOSVMERTVDLPAGTKYVAFRHYNCSDL 1320
 QY 1321 NYLLDDIIOFTMGGSPTDPTVTVYRDGTHKEGLTETTTFFEDGVTATGNHGYCVEVKYT 1380
 DB 1321 NYLLDDIIOFTMGGSPTDPTVTVYRDGTHKEGLTETTTFFEDGVTATGNHGYCVEVKYT 1380
 QY 1381 AGVSPKKVDVTVNSQFNPVQNLTAEOAPNSMDAILKKNAPASKRAEVLNEDFENGIPA 1440
 DB 1381 AGVSPKKVDVTVNSQFNPVQNLTAEOAPNSMDAILKKNAPASKRAEVLNEDFENGIPA 1440
 QY 1441 SWKTIIDADGDGNNWTTTPPPGSSSFAGHNSAICVSSASHINFEQPNPNLYVTPELSPL 1500
 DB 1441 SWKTIIDADGDGNNWTTTPPPGSSSFAGHNSAICVSSASHINFEQPNPNLYVTPELSPL 1500
 QY 1501 GGGTLTFWCAQDANYASEHYAVYASTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
 DB 1501 GGGTLTFWCAQDANYASEHYAVYASTGNDASNFANALLEEVLTAKTVVTAPEAIRGTR 1560
 QY 1561 AQTWTQKTVQLPAGTKYVAFRHFQCTDFFWNLDDVLTSGNAPSVYTTIYRNNTCIAS 1620
 DB 1561 AQTWTQKTVQLPAGTKYVAFRHFQCTDFFWNLDDVLTSGNAPSVYTTIYRNNTCIAS 1620
 QY 1621 GVTEITTRDPLATGFTYGVKVVYPNGESAIEATATNITSLADVTAQKPYTLTVVGKTI 1680

Db 1621 GVTETTYDDPLATGYTYGVKVVYPNGESALETATLNTSLADVTAQKPYTLTVVGKTI 1680
Qy 1681 TVTQGEAMIDMGRRLAAGRNTVVYTAQGGHVAVMVVDGKSVEKLVK 1732
Db 1681 TVTQGEAMIDMGRRLAAGRNTVVYTAQGGHVAVMVVDGKSVEKLVK 1732
RESULT 3
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Revcolde, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Revcolde
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11
Query Match 99.9%; Score 9167; DB 4; Length 1732;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1729; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRKLLLLIAASLLGVLYAOSAKIKLDAPTTRTCTNNSPKQFDASFSNEVELTKVETK 60
Db 1 MRKLLLLIAASLLGVLYAOSAKIKLDAPTTRTCTNNSPKQFDASFSNEVELTKVETK 60
Qy 61 GGTFAVSIPGAPFTGBVGSPEVPAVRKLIAPVGAFTPVVRVKSFTBQVYSLNQYSEKL 120
Db 61 GGTFAVSIPGAPFTGBVGSPEVPAVRKLIAPVGAFTPVVRVKSFTBQVYSLNQYSEKL 120
Qy 121 MPQPSMSKSDDBKVPFVYNAAYARKFGVQBLTOVEMLTMRGVRIAAALTNIPVQYD 180
Db 121 MPQPSMSKSDDBKVPFVYNAAYARKFGVQBLTOVEMLTMRGVRIAAALTNIPVQYD 180
Qy 181 VVANQLKVRNNIEIEVSFGQADEVATQRLYDASPSYPETAYKQLFNRDVVTDHGDLYNT 240
Db 181 VVANQLKVRNNIEIEVSFGQADEVATQRLYDASPSYPETAYKQLFNRDVVTDHGDLYNT 240
Qy 241 FVRMLVYAGAKFKEALKPWLTKAQAQGYLDVHYTDEAEVGTNNASIKAFIHKYNDGLA 300
Db 241 FVRMLVYAGAKFKEALKPWLTKAQAQGYLDVHYTDEAEVGTNNASIKAFIHKYNDGLA 300
Qy 301 ASAPFVLAUVGTDVLSGKGGKTKVKTDLVYSVNDGDYFPEMYTPRMSASSPEELTNI 360
Db 301 ASAPFVLAUVGTDVLSGKGGKTKVKTDLVYSVNDGDYFPEMYTPRMSASSPEELTNI 360
Qy 361 IDKVLMEKATMPDKSLYKLVLLIAGADYSWNSQVQPTIKYGMQYVYNOEHGTVDVNY 420
Db 361 IDKVLMEKATMPDKSLYKLVLLIAGADYSWNSQVQPTIKYGMQYVYNOEHGTVDVNY 420
Qy 421 LKAPYTCYSHLNTGVSFANYTAGHSTAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480
Db 421 LKAPYTCYSHLNTGVSFANYTAGHSTAWADPLLTTSQLKALTNKDKYFLAIGNCCITA 480
Qy 481 QFYDVQPCFGEVITRVEKAGAYIIGSSPNSWGEDYVWSVGNAPVQPTFEGTSMGS 540
Db 481 QFYDVQPCFGEVITRVEKAGAYIIGSSPNSWGEDYVWSVGNAPVQPTFEGTSMGS 540
Qy 541 YDATFLEDSTNTVNSIMWAGNLAATHAGNIGNITHI GAHYWEAYHVLGDGSMVPYRAMP 600

Db 541 YDAIFLEDSTNTVNSIMWAGNLAATHAGNIGNITHI GAHYWEAYHVLGDGSMVPYRAMP 600
Qy 601 KNTYITLIPASLPQNASYSIOASAGSVVAISKDGLYGTGVANASGVATVSMTKOITENG 660
Db 601 KNTYITLIPASLPQNASYSIOASAGSVVAISKDGLYGTGVANASGVATVSMTKOITENG 660
Qy 661 NYDVVITRSNLYLPVIOQVGPSPYOPVSNLTATTOGKVTLKWEAPSAKKAEGSREV 720
Db 661 NYDVVITRSNLYLPVIOQVGPSPYOPVSNLTATTOGKVTLKWEAPSAKKAEGSREV 720
Qy 721 RIGDGLFVTIIEPANDVRANEAKVLAADNVGMDNTGYQFLLDADHNTFGSVIPATGLPFT 780
Db 721 RIGDGLFVTIIEPANDVRANEAKVLAADNVGMDNTGYQFLLDADHNTFGSVIPATGLPFT 780
Qy 781 GTASSNLYSANFEYLVPANADPVTTQNIIVTCGGEVIFGGVYDVCITNPEPASGOWI 840
Db 781 GTASSNLYSANFEYLVPANADPVTTQNIIVTCGGEVIFGGVYDVCITNPEPASGOWI 840
Qy 841 AGDGNOPARYDDFTFEAGKKYFTMERAGMGDGTMEVEDDSDPASVYTYVYRDGTIKE 900
Db 841 AGDGNOPARYDDFTFEAGKKYFTMERAGMGDGTMEVEDDSDPASVYTYVYRDGTIKE 900
Qy 901 GLTATTPEEDGVAAGNHEYCVEVKYTAGVSPKCKDVTVEGSNEFAFPVQNLTCSSVGQKV 960
Db 901 GLTATTPEEDGVAAGNHEYCVEVKYTAGVSPKCKDVTVEGSNEFAFPVQNLTCSSVGQKV 960
Qy 961 TLKWDAPNGTNPENPNPNPGTTLSESFENGIPASWKTIDADGDHGWFGNAPGIAGY 1020
Db 961 TLKWDAPNGTNPENPNPNPGTTLSESFENGIPASWKTIDADGDHGWFGNAPGIAGY 1020
Qy 1021 NSNGCVYSESFGGLGIVLTPDNYLITPALDLPNGGKLTFWCAQDANYASEHVAIVASS 1080
Db 1021 NSNGCVYSESFGGLGIVLTPDNYLITPALDLPNGGKLTFWCAQDANYASEHVAIVASS 1080
Qy 1081 TGDASNFTNALLEETITAKGVSPKAIKRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
Db 1081 TGDASNFTNALLEETITAKGVSPKAIKRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
Qy 1141 FYIDLDEVELKANGKADFTETESSTHGEAPAEWTIDADGGQGWCLCLSSQGLDWLTA 1200
Db 1141 FYIDLDEVELKANGKADFTETESSTHGEAPAEWTIDADGGQGWCLCLSSQGLDWLTA 1200
Qy 1201 HGSNNVVSFWMNGMALNPDNYLISKDVTGATKVYKYAYAVNDGFPGDHVAVMSKGTNA 1260
Db 1201 HGSNNVVSFWMNGMALNPDNYLISKDVTGATKVYKYAYAVNDGFPGDHVAVMSKGTNA 1260
Qy 1261 GDTTVVFEETPNCINKGARFGLSTANGAKPOSVMERTVDLPAGTKYVAFRHYNCSDL 1320
Db 1261 GDTTVVFEETPNCINKGARFGLSTANGAKPOSVMERTVDLPAGTKYVAFRHYNCSDL 1320
Qy 1321 NYILLDDIOFTMGGSPTPTDYTVYVRDGTIKIEGLTETTFEEDGVATGNHGYCDEVKVT 1380
Db 1321 NYILLDDIOFTMGGSPTPTDYTVYVRDGTIKIEGLTETTFEEDGVATGNHGYCDEVKVT 1380
Qy 1381 AGVSPKCCVDVTVNSQFNPVQNLTAEQAPNSMDAILKWNAPASKAEVLNEDFENGIPA 1440
Db 1381 AGVSPKCCVDVTVNSQFNPVQNLTAEQAPNSMDAILKWNAPASKAEVLNEDFENGIPA 1440
Qy 1441 SWKTIDADGDGNWTTTPPGSSFAGHNSAICVSSASHINFEQPONPNYLVTPELSLP 1500
Db 1441 SWKTIDADGDGNWTTTPPGSSFAGHNSAICVSSASHINFEQPONPNYLVTPELSLP 1500
Qy 1501 GGGTLTFWCAQDANYASEHVAIVASSSTNDASNFANALLEEVLTAKTVVTAPAIIRGR 1560
Db 1501 GGGTLTFWCAQDANYASEHVAIVASSSTNDASNFANALLEEVLTAKTVVTAPAIIRGR 1560
Qy 1561 AQTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLDDVWITSNAPSYYTYIYRNNTQIAS 1620
Db 1561 AQTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLDDVWITSNAPSYYTYIYRNNTQIAS 1620
Qy 1621 GVTETTYRDPDLATGFTYGVKVVYPNGESALETATLNTSLADVTAQKPYTLTVVGKTI 1680

| | | | |
|----|------|--|------|
| QY | 1557 | RGTRAQGTWYQKTVOLPAGTKYVAFRHFPGCTDFEWINLDDVVIT-SGNAPSYYTIVIRNN | 1615 |
| Db | 1190 | NLTREK-----DLPAGTKYVAFRHNCTDVLGIMDDWVITGEGGSPYYTVVRDG | 1241 |
| QY | 1516 | TQTASGYTETTYRPPDLATGYTVGVKVPVNGESALETATLNTLSADYTAQKPYLTIV | 1675 |
| Db | 1242 | TKIQEGLTETTYRADGMSAOSHESYCEVFKYAAGVSPKVCVDYIPDGVAADVTAQKPYLTIV | 1301 |
| QY | 1676 | VGKTLITVTCGEAMIYDMNGRRLAAGNTVVYTAQGGHYAMVVVDGKSVTEKLAVK | 1732 |
| Db | 1302 | VGKTLITVTCGEAMIYDMNGRRLAAGNTVVYTAQGGYAMVVVDGKSVTEKLAIK | 1358 |

RESULT 5
 US-08-141-324-14
 ; Sequence 14, Application US/08141324
 ; Patent No. 5475097
 ; GENERAL INFORMATION:
 ; APPLICANT: Travis, James
 ; APPLICANT: Potempa, Jan S.
 ; APPLICANT: Barr, Philip J.
 ; APPLICANT: Pavloff, Nadine
 ; APPLICANT: Pike, Robert N.
 ; TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis
 ; TITLE OF INVENTION: Protease
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee and Winner, P.C.
 ; STREET: 5370 Manhattan Circle, Suite 201
 ; CITY: Boulder
 ; STATE: CO
 ; COUNTRY: US
 ; ZIP: 80303

| | Query Match | 52.4% | Score 4811; | DB 1; | Length 942; |
|----|-----------------------|---|----------------|-----------|-------------|
| | Best Local Similarity | 96.7% | Pred. No. 0; | | |
| | Matches 911; | Conservative 17; | Mismatches 14; | Indels 0; | Gaps 0; |
| QY | 1 | MRKLLLLTAASLLGVGLYAQSAKILKDAPTRTRCTNNSPKQPDASFSFNEVELTKVETK | 60 | | |
| DB | 1 | MRKLLLLTAASLLGVGLYAQNAKILKDAPTRTRCTNNSPKQPDASFSFNEVELTKVETK | 60 | | |
| QY | 61 | GCTFASVSIPGAFTTGEVGSPEVPAVRKLIAPVGATPVVRVKSFTFQVYSLNQVGSSEKL | 120 | | |
| DB | 61 | GCTFASVSIPGAFTTGEVGSPEVPAVRKLIAPVGATPVVRVKSFTFQVYSLNQVGSSEKL | 120 | | |
| QY | 121 | MPHQESMSKSDDPKVPFVYNNAAAYARKGVFGELTQVEMLTGMRGVRITAAALTINPVOYD | 180 | | |
| DB | 121 | MPHQESMSKSDDPKVPFAYNAAAYARKGVFGELTQVEMLTGMRGVRITAAALTINPVOYD | 180 | | |

| | | | |
|--|-----|--|-----|
| QY | 181 | VVANQLKVRNNIEIEVVSFCQADEVATORLYDASFSPYFETAYKQLFNRRDYTTDGHGLYNT | 240 |
| DB | 181 | VVANQLKVRNNIEIEVVSFCQADEVATORLYDASFSPYFETAYKQLFNRRDYTTDGHGLYNT | 240 |
| QY | 241 | PVRMLVAVAGAKPEKALPKWLTKWAKGFLVDVHYTDEAEVGTTNASIKAFIHKKYNDGLA | 300 |
| DB | 241 | PVRMLVAVAGAKPEKALPKWLTKWAKGFLVDVHYTDEAEVGTTNASIKAFIHKKYNDGLA | 300 |
| QY | 301 | ASAAPVFLALVGDGTDVIISSEKGGKTKKVTDLIYSYSAVDGDYFPFEMTFRMSASSPEELTNI | 360 |
| DB | 301 | ASAAPVFLALVGDGTDVIISSEKGGKTKKVTDLIYSYSAVDGDYFPFEMTFRMSASSPEELTNI | 360 |
| QY | 361 | IDKULMYEKATMPDKSYLEKVLIIAGADYSWNNSVQVQPTIKYGMQYYYNQEHGYTDDVNY | 420 |
| DB | 361 | IDKULMYEKATMPDKSYLEKVLIIAGADYSWNNSVQVQPTIKYGMQYYYNQEHGYTDDVNY | 420 |
| QY | 421 | LKAPYTCGVSHLNTGVSFANTYAHGSETAWADPLATTSOLKALTNKKKYFLAIGNCCIATA | 480 |
| DB | 421 | LKAPYTCGVSHLNTGVSFANTYAHGSETAWADPLATTSOLKALTNKKKYFLAIGNCCIATA | 480 |
| QY | 481 | QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYIYSVGANAVFGVQPTFEGTSMGS | 540 |
| DB | 481 | QFDYVQPCFGEVITRVKEKGAYAYIGSSPNSYWGEDYIYSVGANAVFGVQPTFEGTSMGS | 540 |
| QY | 541 | YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGHAYTWEAYHVLGDGSMVMPYRAMP | 600 |
| DB | 541 | YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGHAYTWEAYHVLGDGSMVMPYRAMP | 600 |
| QY | 601 | KTNVTTLPASLPQOQASYSIOASAGSYVAISKDGLVYGTGVANASGVATVMTKQITENG | 660 |
| DB | 601 | KTNVTTLPASLPQOQASYSIOASAGSYVAISKDGLVYGTGVANASGVATVMTKQITENG | 660 |
| QY | 661 | NYDVVITRSNLYPVIKQIQVGEPSPYQPVSNLTATTOGQKVTLKWEAPSAKKAEGSREVK | 720 |
| DB | 661 | NYDVVITRSNLYPVIKQIQVGEPSPYQPVSNLTATTOGQKVTLKWEAPSAKKAEGSREVK | 720 |
| QY | 721 | RIGDGLFVTIPANDVRANEAKVLAAADNVGMDNTGYOFLDADHNTFGSVIPATGPLFT | 780 |
| DB | 721 | RIGDGLFVTIPANDVRANEAKVLAAADNVGMDNTGYOFLDADHNTFGSVIPATGPLFT | 780 |
| QY | 781 | GTASSNLYSANFEYLVPANADPVVTTQNIIVTGOGEVVI PGVVDYCYITNPEPASGKMWI | 840 |
| DB | 781 | GTASSNLYSANFEYLVPANADPVVTTQNIIVTGOGEVVI PGVVDYCYITNPEPASGKMWI | 840 |
| QY | 841 | AGDGGNQARYDDPTTFBAGKYYITMRAGNGDGTDMEDDDSPASYTYTVYVRDGTIKE | 900 |
| DB | 841 | AGDGGNQARYDDPTTFBAGKYYITMRAGNGDGTDMEDDDSPASYTYTVYVRDGTIKE | 900 |
| QY | 901 | GLTATTFEEDGVAAGNHEYCVVEKYTAGVSPKVKDVTVEGS | 942 |
| DB | 901 | GLTATTFEEDGVAAGNHEYCVVEKYTAGVSPKVKDVTVEGS | 942 |
| RESULT 6 | | | |
| US-08-541-902-14 | | | |
| ; Sequence 14, Application US/08541902 | | | |
| ; Patent No. 5707620 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Travis, James | | | |
| ; APPLICANT: Potempa, Jan S. | | | |
| ; APPLICANT: Barr, Philip J. | | | |
| ; APPLICANT: Pavloff, Nadine | | | |
| ; APPLICANT: Pike, Robert N. | | | |
| ; TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis | | | |
| ; TITLE OF INVENTION: Protease | | | |
| ; NUMBER OF SEQUENCES: 28 | | | |
| ; CORRESPONDENCE ADDRESS: | | | |
| ; ADDRESSEE: Greenlee and Winner, P.C. | | | |
| ; STREET: 5370 Manhattan Circle, Suite 201 | | | |
| ; CITY: Boulder | | | |
| ; STATE: CO | | | |
| ; COUNTRY: US | | | |

RESULT 6
US-08-541-902-14
; Sequence 14, Application US/08541902
; Patent No. 5707620
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; APPLICANT: Pike, Robert N.
; TITLE OF INVENTION: Lysine-specific Porphyromonas gingivalis
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US

| | | | |
|---|----|---|-----|
| 601 | Db | KTNTYTLPSLFPQNASYSIQASAGSYVAISKDGLVIGTVANAGSVATVNNKQITENG | 660 |
| 661 | Qy | NYDVWITRSNLYPVKIQIVGSPSPYQPVSNLTATQGOVKTLKWEAFSAKKAEGSRVK | 720 |
| 661 | Db | NYDVWITRSNLYPVKIQIVGSPSPYQPVSNLTATQGOVKTLKWDAPSAKKAEGSRVK | 720 |
| 721 | Qy | RIGDGLFVITIEPANDVRANEAKVLAADNVGDNITGYQFLLDADINTGSIPIATGPIFT | 780 |
| 721 | Db | RIGDGLFVITIEPANDVRANEAKVLAADNVGDNITGYQFLLDADINTGSIPIATGPIFT | 780 |
| 781 | Qy | GTASSNLYSANFEYLVPANADPVWTTQNIIVTGOGEVWIPGVGVYDCTINPEPASGKWI | 840 |
| 781 | Db | GTASSNLYSANFEYLVPANADPVWTTQNIIVTGOGEVWIPGVGVYDCTINPEPASGKWI | 840 |
| 841 | Qy | AGDGNQPARYDDFTFEAGKKTFTMRAGMDGDTMEVEDDPSASYTYTVYRDGTIKE | 900 |
| 841 | Db | AGDGNQPARYDDFTFEAGKKTFTMRAGMDGDTMEVEDDPSASYTYTVYRDGTIKE | 900 |
| 901 | Qy | GLTATTFEEDGVAAGNHCEYCVKVTAGVSPKVKDVTVEGS | 942 |
| 901 | Db | GLTATTFEEDGVAAGNHCEYCVKVTAGVSPKVKDVTVEGS | 942 |
| RESULT 7 | | | |
| US-08-570-311-8 | | | |
| ; Sequence 8, Application US/08570311 | | | |
| ; Patent No. 5824791 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Proguiske-Fox, Ann | | | |
| ; APPLICANT: Tumwasorn, Somying | | | |
| ; APPLICANT: Lepine, Guylaine | | | |
| ; APPLICANT: Han, Naiming | | | |
| ; APPLICANT: Lantz, Marilyn | | | |
| ; APPLICANT: Patti, Joseph | | | |
| ; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes | | | |
| ; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease | | | |
| ; NUMBER OF SEQUENCES: 29 | | | |
| ; CORRESPONDENCE ADDRESS: | | | |
| ; ADDRESSEE: Ted W. Whitlock | | | |
| ; STREET: 2421 N.W. 41st Street, Suite A-1 | | | |
| ; CITY: Gainesville | | | |
| ; STATE: FL | | | |
| ; COUNTRY: USA | | | |
| ; ZIP: 32606 | | | |
| ; COMPUTER READABLE FORM: | | | |
| ; MEDIUM TYPE: Floppy disk | | | |
| ; COMPUTER: IBM PC compatible | | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| ; SOFTWARE: PatentIn Release #1.0, Version #1.25 | | | |
| ; CURRENT APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US/08/570,311 | | | |
| ; FILING DATE: | | | |
| ; CLASSIFICATION: 424 | | | |
| ; PRIOR APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US 08/353,485 | | | |
| ; FILING DATE: 09-DEC-1994 | | | |
| ; CLASSIFICATION: 424 | | | |
| ; PRIOR APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US 07/647,119 | | | |
| ; FILING DATE: 25-JAN-1991 | | | |
| ; CLASSIFICATION: 424 | | | |
| ; PRIOR APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US 07/241,640 | | | |
| ; FILING DATE: 08-SEP-1986 | | | |
| ; ATTORNEY/AGENT INFORMATION: | | | |
| ; NAME: Whitlock, Ted W. | | | |
| ; REGISTRATION NUMBER: 36,965 | | | |
| ; REFERENCE/DOCKET NUMBER: UP15.C3 | | | |
| ; TELECOMMUNICATION INFORMATION: | | | |
| ; TELEPHONE: (904) 375-8100 | | | |
| ; TELEFAX: (904) 372-5800 | | | |
| ; INFORMATION FOR SEQ ID NO: 8: | | | |

Query Match 50.9%; Score 4673.5; DB 2; Length 1087;
Best Local Similarity 81.4%; Pred. No. 1e-312;
Matches 900; Conservative 56; Mismatches 120; Indels 29; Gaps 10;

QY 638 CTGVANASGATVSMTKITENGYNVDWITRSNYLFPVKIQIGSPSPQPVSNLTATQQ 697
||| : : |||

Db 2 GTVVADPTTAAFPVKAKQIAENGNDVVMTRSNLYLFINQIQAGEPSPYQPVNLTAPPE 61
Qy 698 GOKVTLKWEASAKKAEGSREVKRIIGDLGFLVITIEPANDVRANEAKVLAADNMVGDNTGY 757
Db 62 GEEVAKWDTTSAKKAERAKSREVKRIIGDLGFLVITIEPANDVRANEAKVLAADNMVGDNTGY 121
Qy 758 QFLDADHNTGSGVIPATGPIPLTGTASSNLXANFEYLVPANADPVVTTQNIIVTGOGEV 817
Db 122 QFLDADHNTGSGVIPATGPIPLTGTASSNLXANFEYLVPANADPVVTTQNIIVTGOGEV 181
Qy 818 VIPGVYDYCYTNPPEPAGKMWIADGNGQAPARYDDFTFEAGKKYFTMRAGMGDGTDM 877
Db 182 VIPGVYDYCYTNPPEPAGKMWIADGNGQAPARYDDFTFEAGKKYFTMRAGMGDGTDM 241
Qy 878 EVEDDSPASYTIVVRDGTKEGLTATTFEEDGVAAGNHEVCVEVKYTAGVSPKVKCDV 937
Db 242 EVEDDSPASYTIVVRDGTKEGLTATTFEEDGVAAGNHEVCVEVKYTAGVSPKVKCDV 301
Qy 938 TVESNEFAPVQNTLGSSVGQKVTILKWDAPNGTNPENPNPNTGTTLSSEFENGIPASW 997
Db 302 TVESNEFAPVQNTLGSSVGQKVTILKWDAPNGTNPENPNPNTGTTLSSEFENGIPASW 360
Qy 998 KTIADGCHGWKPGNAPGIAAGNSGCVYSESFGIGIGIVLTPDNYLITPALDLPNGK 1057
Db 361 KTIADGCHGWKPGNAPGIAAGNSGCVYSESFGIGIGIVLTPDNYLITPALDLPNGK 420
Qy 1058 LTFWVCAQDANYASEHAYVAYASSTNDASNTNALLLETTITAKGVRSKPAIRGRIOGTWR 1117
Db 421 LTFWVCAQDANYASEHAYVAYASSTNDASNTNALLLETTITAKGVRSKPAIRGRIOGTWR 480
Qy 1118 QKTVDLPAGTKYVAPRHPQSTDMFYIDLDEVEIKANGKRAADFTTETFESSHGEAPAEWTT 1177
Db 481 QKTVDLPAGTKYVAPRHPQSTDMFYIDLDEVEIKANGKRAADFTTETFESSHGEAPAEWTT 540
Qy 1178 IDADGCGGOWCLSSGQDLWLTARGGNSVVSFSWNGMALPNDNYLISKDVTGATKVKY 1237
Db 541 IDADGCGGOWCLSSGQDLWLTARGGNSVVSFSWNGMALPNDNYLISKDVTGATKVKY 600
Qy 1238 YAVNDGPGDHAYVAMI SKTGNAGDFTVVFETPNKNGGARGLSTPEANGAKPOSVWI 1297
Db 601 YAVNDGPGDHAYVAMI SKTGNAGDFTVVFETPNKNGGARGLSTPEANGAKPOSVWI 660
Qy 1298 ERTVDLPAGTKYVAPRHPYNSDLNLYLDDIQTMTGGSPTTDTYTYVVRDGTIKKEGLT 1357
Db 661 ERTVDLPAGTKYVAPRHPYNSDLNLYLDDIQTMTGGSPTTDTYTYVVRDGTIKKEGLT 720
Qy 1358 ETTTEEDGVAAGNHEVCVEVKYTAGVSPKCVDTVNSTQNPQNLTAEQAPNSMAIL 1417
Db 721 ETTTEEDGVAAGNHEVCVEVKYTAGVSPKCVDTVNSTQNPQNLTAEQAPNSMAIL 778
Qy 1418 KWNAPASKRAEVLNEDFE-NGIPASWKITIDADGDNW- -TTTPPPG- -SSFAGN 1469
Db 779 KWEAPSGKRGELLNEDFEADIPGTWALDADGNGNWDITLNEFTGERHVLSPLRASN 838
Qy 1470 SAICVSSASHINFGQPNPNLYLTPELSLPGGTLTFWCAQD- -ANYASEHAYVAYASST 1528
Db 839 VAISYSLLOQOEYLPNPNFLITPKVE- -GAKKITVKGSGPLQWMSHDHYALCISKS 896
Qy 1529 GNDASNFANALLBEVLTAKTIVTAPAIRGTRAGQVTKVQVLPAGTKYVAFRHFCTD 1588
Db 897 GTAAADF- -SVIPEETMTYTTQGANLPREK- -DLPAGTKYVAFRHYNCTD 942
Qy 1589 FFWNLNDDVIT- -SGNAPSYYTYTYRNNTQIASGVTTETTYRDPDLATGYTYGVKAVYPN 1647
Db 943 VLGMIDDDVITGEGEGPSYTYTYVRDGTKEGLTETTYRDAQMSAQSHCYVEVKIAA 1002
Qy 1648 GESAIETATLNTSLADVTAQKPYTLTVVGKTIITVTCQGEAMIYDMNGRRLAAGNTVY 1707
Db 1003 GVS PKVCVDYIPDGVADVTAQKPYTLTVVGKTIITVTCQGEAMIYDMNGRRLAAGNTVY 1062
Qy 1708 TAQGGYAVMVVVDGKSYVEKLVAK 1732
Db 1063 TAQGGYAVMVVVDGKSYVEKLVAK 1087

RESULT 9

US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666

; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada

; FILE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE

; CURRENT APPLICATION NUMBER: US/09/066,330A

; CURRENT FILING DATE: 1998-09-15

; EARLIER FILING DATE: 1995-10-30

; EARLIER APPLICATION NUMBER: PCT/AU96/00673

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1706

; TYPE: PRT

; ORGANISM: Porphyromonas gingivalis

US-09-066-330-10

Query Match 46.5%; Score 4272; DB 4; Length 1706;
Best Local Similarity 50.9%; Pred. No. 8.9e-285;
Matches 932; Conservative 204; Mismatches 456; Indels 240; Gaps 47;

Qy 6 LLIAASLLGLVLAQSAKIKLADPTTCTTNNNGFKOFDASFSFNEVELTKVETKGGTFA 65
Db 10 IALCSSLLGGMAFAQQTELGRNPNVRLLESQQSVTK- -VQFRMDNLKFTVQTPKGI- - 65
Qy 66 SVSIPGAPPT- - - - -GEVSGPEVPAVKLIAPVPGATPVVRYS- -FTEQVYSLNQYG 116
Db 66 - - - - -GQVPTVTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVVSKEFKKNVL- - - - 116
Qy 117 SEKLMPHQPMSKSDDEKPEKVPVNAAYARKGVGQELTOVEMGLTMRGVRIALTTNP 176
Db 117 - - - - -IAPSKGIMRNEDPKPIPVY- -GKTSQNKFPGEIATLDDPFIILDRVGRQVWNPAP 172
Qy 177 VOYDVANQLKVRNNIEI- - -EVSFGQADEVATQRLVDASFPYFETAYKOLFNRDVYTD 233
Db 173 LQYNPVTYTLRIYETIIVAVSETSEQKNILKKGTFAG- - - - -FETYKRMF- - - - -MN 222
Qy 234 HGDLYNTPV- - - - -RMLVWAGAKFKKALPWLTKAQKGFYLDVHYTDEAEVGTNASI 287
Db 223 YEPGRYTPVEEKQNGRMIVIVAKKVEGDIKDFVDKMQNQLRTEVKVVAEDIASPTANAI 282
Qy 288 KAFTHKXY- - - - -NDGLAASAPVFLALVGDTDVTSGB- -KGGTKKVTDLIYSAVDGDYF 341
Db 283 QQFYKQYKEGND- - - - -LTYVLLIGHDKDIPAKITPGIKSDQV- - - - -YGQIVGNDHY 332
Qy 342 PEMYTERMSASSPEELTNIIDKLMVEKATMPDKSYLEKVLIIAGADYSWNISQVQPTIK 401
Db 333 NEVFIGFESKEDLTKQIDRIIHERNTTETDKMLQALCIASAGGSGSADGESDIO 392
Qy 402 Y-GMQYYNNOEHGYTDVYNLYKAPYTCY- - - - -SHLNTGVSFANYTAGSETAW 450
Db 393 HENVIANLLTQYGYTKIK- - - - -CYDPGVTPKNIIDAFNGGISLANVTGSETAW 444
Qy 451 ADPLLTTSQALKATNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV- - - - -KEKGAYAYIG 506
Db 445 GTSHFGTHVKQNTNSQLPEIFDVACVNGDFLFSMPCFAEALMRAQKQKGTGTVAIIA 504
Qy 507 SSPNSWGEDYVYWSGVANAVFGVQPTFEGTSMGSDATFLEDSYNTVNSIMWAGNLAATH 566
Db 505 STINQSW- - - - -ASPMRG- - - - -QDEMNEI- - - - -LCEKH 529
Qy 567 AGNI- - - - -GNITHGAHYWEAYH- - - - -VLGDGVSMPYRAMPKNTIYVTPASL 611
Db 530 PNNIKRTFGVTVWNGMPAMVEKYKDGKMKLDTWTVFGDPSLLVRLTVPTKMQVTAQAI 589

177 VOYDVVANOLKVRNTEI---EVSFOGADEAVATQRLYDASFSFYFETAYKQLENRDVYTD 233
173 LQNPVTKLRIYETITVAVSETSEBQKNLNLKGGTFAG-----FEDTYKRMF-----MN 222
234 HGLDYNTPV-----RMLVVAGAKFKALKPWLTKAOKGFYLDVHYTDEAEVGTITNASI 287
223 YEPGRVTPVEEKQNGRMIVIVAKKYEKGDJKDFVDMKNQGLRTEVKVAEDIASPVTANAI 282
288 KAFIHKY-----NDGLAASAPFLALVGTDVISGE--KGKTKKQVTDLDIYSAVDGYF 341
283 QQFVKQYEYKEGND-----LTYVLLVGHDKDIPAKITPGIKSDQV---YQIIVGNDHY 332
342 PEMYTERMSASSPEELTNIIDKVLMEKATMPKSVLEKVLVLLAGADYSWNSQVQPTIK 401
333 NEVFIGRCESEKEDUKTQIDRIIHERNITTEDKWLQGLCALCIASAEGBGSADNGESDIQ 392
402 Y-GMOYVYNQEBGYTDVYNLKAFTYGCY-----SHLTGVSFANYTAHGETAW 450
393 HENVIANLLTQYGYTKIILK-----CYDPGVTPKNIIDAFNGGILSVNYTCHGSETAW 444
451 APBLTTSOLKALTNKDKYFLATGNCICIITQAFYVOPCFGEVITRV-----KEKGAVAYIG 506
445 GTHSGFTTHVKQLTNSGNOLPFIIDVACVNGDFFLSMPCFAEALMRAQKDGKPTGTVAIIA 504
507 SGPNSYWGEDIYWSVGANAVFGVQPTFEGTSMGSIYDATFLEDSYNTVNSIMWAGNLAATH 566
505 STINQSW-----ASPMRG-----QDEMNEI-----ICEKH 529
567 AGNI-----GNITHIGAHYWEAYH-----VLGDSVMPYRAMPKNTYTLPSAL 611
530 PNNIKRTFGVTMNGMFAMVEKYKDKGEKMLDFTWTFVGDPSLLVRLVPTKMQVTAPAQI 589
612 PQWQASYSYQAS--AGSYVAISKDGLVYGTGVANASGVATVSMTKQITENGYNVIVITRSN 670
590 NLTDASVWNSCDYNGAIAITISANGKMFSGSAVBE--NGTATINLT--GLTNBETSLTLTVVGYN 647
671 YLPVIKQIQV--GEPSPYQPVNSLTATTQOKVTLKWEAPSAK---KAEGSREVKRIGDGL 726
648 KETVIKTINGEPNPYQPVNSLTATTQOKVTLKWDAPSTKTNATTNARSVDGIRELV 707
727 FVITIEPAND--VRANEAKVLAADNVMDNTGYQFLLDADHNTFGSVIPA--TGPLFTG--TA 783
708 LLSVSDAPELLRSGQAEIVLEAHADVNDGSGYQILLDADHDQXGQVIPSHTHLWENCVS 767
784 SSNLYSANFEYLIPANADPVWITQNIIVTCQGVWIPGGVYDYCIINPEPASGKMIAGD 843
768 PANLF--APPEYTVPENADPSCSPTNIMDGTASVNIIPAGTYDFAIAAPQ--ANAKIWIAGQ 825
844 GGNQPARYDDFTPEAGKKTFTMRRAGMGDGTDMVEDEDDSPASYTYTVYRDGFKIKEGLT 903
826 G---PTKEDDYVEAGKKYHFLMKMGSGDGTBLTISEGGSDYTYTVYRDGFKIKEGLT 882
904 ATTFEEDGVAAGNHEVCVEVKYTAGVSPKCKDVTVEGSNEFAPVQNLTGSSYGQKVTLK 963
883 ATTFEEDGVAAGNHEVCVEVKYTAGVSPKCKDVTVEGSNEFAPVQNLTGSAVGQKVTLK 942
964 WDPNGTTPNPNPNPNPG--TTLISEFENGIPASWKTIDADGCHGKPGNAPGIAGYNS 1022
943 WDPNGTTPNPNPNPNPGTTLISEFENGIPASWKTIDADGCHGKPGNAPGIAGYNS 1002
1023 NGCVYSESFLGIGIGVLTDPNYLITPDLNPGKGLTFWVCAODANVASHYAVYASSTG 1082
1003 NGCVYSESFLGIGIGVLTDPNYLITPDLNPGKGLTFWVCAODANVASHYAVYASSTG 1062
1083 NDASNFTNALLEETITAKGVRSPKAIIRIGTQWROKTVLDLPAGTKYVAFRHFQSTDMFY 1142
1063 NDASNFTNALLEETITAKGVRSPKAIIRIGTQWROKTVLDLPAGTKYVAFRHFQSTDMFY 1122
1143 IDLDEVEIKANGKRAPDTETFEESTHGEAPAEWTTIDADGGQGWLCSSGQLDWLTAGH 1202
1123 IDLDEVEIKANGKRAPDTETFEESTHGEAPAEWTTIDADGGQGWLCSSGQLDWLTAGH 1182

1203 GSNVYSSFSWNGMALNPDNVLISKQVGTGATKVKYVAVVNDGPPGDHYAVMISKTGTNAGD 1262
1183 GINVASFSWNGMALNPDNVLISKQVGTGATKVKYVAVVNDGPPGDHYAVMISKTGTNAGD 1242
1263 FTWFEETPNNGKNGKARFGLSTEANGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLYN 1322
1243 FTWFEETPNNGKNGKARFGLSTEANGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLYN 1302
1323 ILLDDIOFTMGSGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVAATGNHEVCVEVKYTAG 1382
1303 ILLDDIOFTMGSGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVAATGNHEVCVEVKYTAG 1362
1383 VSPKCKVDVTVNSTQFNPVQNLTAEOAPNSMDALLKWNAPASKRAE-----1428
1363 VSPKCVNVTIINPTQNPVZNLKAQ---PDGGDVVLKWEAPSAKKTGESSREVKRIKIGDLFV 1420
1429 -----VLNEDFENGIPASWKTIDADGDNNTWTTTPPPGGSSFAGHNSAIC 1473
1421 TIIEPANDVRANEAKVLAADNVMDNTGYQFL--LDADHNTFGSVIPATGPTGTASSNL 1479
1474 VSSASHINPEG--PQNPD-----NYLVT--BELSLPGGGLTTFWVCAODANVASHYAV 1523
1480 YSA---NFEYLIPANADPVWITQNIIVTCQGEVVIPIGG---VIDYCIINPEPASGKMI 1532
1524 YASSTGNDASNFANALLEEVLTAKTVVTAPALRGTRAQGTWYQKTVQLPAGTKYV--AFR 1582
1533 -AGDGNQPARVDDFTFE-----AGKKTFTWTR 1559
1583 HFGCTDFFWINLDDVVITSGNAPSYYTYTYIRNNTQIASGVTTETTYRDPDLATGYTYGVK 1642
1560 RAGMGDG---TDMVEDEDDSPASYTYTVYRDGTKIKEGLTETTYRDAAGSAQSHEVCVE 1614
1643 VYVPNGESALETATLNTISLADVTAQKPYTLTVVGTITVTCQGEAMIVDMNGRRLAAGR 1702
1615 VKYAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGTITVTCQGEAMIVDMNGRRLAAGR 1674
1703 NTWYTAQGHYAVVWVVDGKSVVEKLAVK 1732
1675 NTWYTAQGHYAVVWVVDGKSVVEKLAVK 1704

RESULT 11

US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917
; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.
 REGISTRATION NUMBER: 33,878
 REFERENCE/DOCKET NUMBER: 103-95 WO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 488-8080
 TELEFAX: (303) 499-8089
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1704 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-822-324-6

Query Match 46.5%; Score 4270; DB 3; Length 1704;
 Best Local Similarity 50.9%; Pred. No. 1.2e-284;
 Matches 932; Conservative 204; Mismatches 456; Indels 238; Gaps 47;

```

QY 6 LLIAASLLGLVLYAQSAAKIKLDAPTTRTCTNNFKQPDASFSENEVELTKVETKGGTFA 65
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 10 IALCSSLLGMAFAQQTELGRNPNVRLLESTQQSVTK--VQFRMDNLKFTVQTPKGI-- 65
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 SVSIPGAPPT-----GEVSGSEVPVAKLIAPVPGATPVVYKS--FTQVYSLNQYG 116
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 -----GQVPTTEGVNLSSEKMPILILSRSLAVSDTREMKVEVSVSKFIEKKNVL---- 116
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 SEKLPHQPSMKSDDPEKVPFVYNAAYARKGFVGBELTOVEMLGTMRGVRIAAATINP 176
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 ---IAPSGMIMRNEDPKIFVYV--GKSYSQNKFPFGGIAILDOPFILRDVRGQVNPAP 172
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 VOYDVANQLKVRNRIE---EVFQGADEVATORLYDASPSFETAYKOLFNRDVTVD 233
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 LQYNPVTILTYITVAVSETSEQGNILNKKGTAG----FEDTYKRMF-----MN 222
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 HGDLYNTPV-----RMLVAGAKFKEALKEWLTWKAQKQFYLVDVHYDEAEVGTNASI 287
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 YEPGRYIPVESEKQGRMILVIYAKYEGDIKDFVDMKNORGLRTEVKVAEDIASPVANAI 282
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 KAFIHKY-----NDGLAASAPFVIALVGDGTDVISE--KGKTKKVTDLVYSVADGDYF 341
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 QQFVKQYEKEGND-----LTVLLVGDHDKIPAKITPGIKSDQV--YQIVGNDHY 332
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 PEMYTFMSASPELTNIIIDKVLMEKATPDKSYLEKVLIIAGADYSWNSQVQPTIK 401
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 NEVFIGRSCSEKDELTKQIDRTIHYERNITTEDKWLQALCIASAEGPSADNGESDIQ 392
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 Y-GMYYNNQHGHTDVYNNYLKAPYTCY-----SHLNTGVSPANTYANGSETAW 450
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 393 HENVIANLLTCYGYTKIIL-----CYDPGVTPKNIIDAFNGISLVNTYHGSETAW 444
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 451 ADPLLTTSQKALTNKQYFLAIGNCCITAOFDYVQPCFGEVITRV-----KEGAYAYIG 506
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 445 GTSFHGTTHVKQLTNSQLPFIDVACVNGDFLFSMPCFAEALMBAQDKGKPTGVALLA 504
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 507 SSPNSYGEDYVWSGANAVGVQPTFGTSGMSGDATFLEDSYNTVNSIMWAGNLAATH 566
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 505 STINQSW-----ASPMRG-----QDEMNEI-----LCEKH 529
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 567 AGNI-----GNITHGAHYWEAYH-----VLGDGSMVPYRAMPKNTVTLPASL 611
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 530 PNNIKRTGGVTNGMFAWVEKYKDGKEMLDLTWTFGDPILLVRLVPTPKQVTAQAQI 589
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 612 PQNQASYSIQAS-AGSYVAISKDGLVLYGTGVANASGVATVSMTKQITBNGYDVVITRSN 670
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 590 NLTDASVNVSCDYNGAIAIATISANGMFGSAVVE-NGTATINLT-GLTNESLTTLTVVGVN 647
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 671 YLPVTKIQV-GEPSYQVSNLTATTOGKVLKWEAPSAK---KAGSREVRIRIGDGL 726
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 648 KETVIKINTNGERNPYQVSNLTATTOGKVLKWEAPSAK---KAGSREVRIRIGDGL 707
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 727 FVTIEPAND-VRANEAKVVLAAADNVWGDNTGYQFL-LLADHNTFGSVIPATGLFTTASSNL 1479
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 708 LLSVSDAPELLRSQAIEVLEAHVDWNDGSGYQLILLADHDHQQYQGVIPSDHTLWPNCSV 767
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12

US-09-490-931-10

; Sequence 10, Application US/09490931

Patent No. 6274718
 GENERAL INFORMATION:
 APPLICANT: Travis, James
 APPLICANT: Potempa, Jan S.
 APPLICANT: Barr, Philip J.
 APPLICANT: Pavloff, Nadine
 TITLE OF INVENTION: Porphyromonas gingivalis
 TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 STREET: 5370 Manhattan Circle, Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: US
 ZIP: 80303

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/490,931
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/336,308
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/265,441
 FILING DATE: 24-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferber, Donna M.
 REGISTRATION NUMBER: 33,878
 REFERENCE/DOCKET NUMBER: 21-93C
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 499-8080
 TELEFAX: (303) 499-8089
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1704 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-490-931-10

Query Match 46.5%; Score 4270; DB 3; Length 1704;
 Best Local Similarity 50.9%; Pred. No. 1.2e-284;
 Matches 932; Conservative 204; Mismatches 456; Indels 238; Gaps 47;
 6 LATAASLLGVLVAQSAKIKLDAPTTRTTCTNNSFKQFDASFNEVELTKVETKGTFA 65
 10 IALCSSLLGMAFAQOTELGRNPNVRLLESTQOSVTK--VQFRMDNLKFTVEVQTPKGI-- 65
 66 SVSIPGAPFT-----GEVGSPEVPAVKLIAPVPGATPVVRVKS--FTEQVYSLNQY 116
 66 ----GQVPTYTEGVNLSEKGMPTLILSRSLAVSDTREMKVEVWSSKFTKKNVL---- 116
 117 SEKLMPHQSMKSDPEKPFVYNAAYARKFVQQLTQVEMLGTMRGVRITAAITNP 176
 117 ---IASKGMINDEPKPIPYV--GKSYSONKFFPFGEIATLDDPFILRDVRGQVNFAP 172
 177 VOYDVVANQLKVRNNIEI---EVSFOGADEVATQRLYDASFSPYFETAYKOLENRDVTYD 233
 173 LQINPTVTLRIYETITVAIVSEIQQKNILNKGTFAG-----FEDTYKRMF-----MN 222
 234 HGLDYNTPV-----RMLVVAGAKFKALKPMLTWAKQGFYLDVHYTDEAEVGTNASI 287
 223 YEPGRYTPVEEKONGRMIVIVAKKYEGDIKDFVDWKNQGRLRTEVVAEDIASPVVANA 282
 288 KAFIHKKY-----NDGLAASNAAPVFLAIVGDTDVISE--KGKTKKQWLDLYSAVDGDF 341
 283 QQEVKQYEKEGND-----LTYVLLVGHDKDIPAKITPGIKSDQV---YGOIVGNHDY 332

QY 342 PEMYTRMSASSPEELTNIIDKVMTEKATMPDKSYLEKVLIIAGADYSWNSQVQPTIK 401
 Db 333 NEVFIGRFSCEKEDLKTQIDRTIHYERNITTEDKWLQALCIAASABGGPSADNGESDIQ 392
 QY 402 Y-GWQYVYNQEHGYTDVYVYLKAPYTCY-----SHLNTGVSFANVTAGSSTAW 450
 Db 393 HENVIANLLTQYGTIKLIK-----CYDPGVTPKNIIDAFNGGISLVNVTGHGSSTAW 444
 QY 451 ADPLLTTSOLKALTNKDKYFLAIGNCCTIAQFDYVQVPCFGEVITRV-----KEKGAVAYIG 506
 Db 445 GTSHFGTTHVKQLTNSQLPFIQFDVACVNGDFLPSMPCFAEALWRAQDKCKPTGTVAIIA 504
 QY 507 SSPNSYWGEDIYWSVGANAVFGVQPTFEGTSMGSDATFLEDSTNTVNSIMWAGNLAATH 566
 Db 505 STINQSGW-----ASPMRG-----QDEMNET-----LCEKH 529
 QY 567 AGNI-----GNITHGAHYWEAYH-----VLGDSVMPYRAMPKTNITYTLPASL 611
 Db 530 PNNIKRTFGVTVNGVFMFAMVEKIKKDGKMLDWTWVFGDSSLVRLTLVPTQMVTPAQI 589
 QY 612 PQOASYSIOAS--AGSYVAISKDGLVYGTGVANASGVATVSMTKQITENGNYDVWITRSN 670
 Db 590 NLTDASVNVSCDYNGAIAIATISANGKMFSGSAVB-NGTATINLT-GLTNESTLTLTVVGYN 647
 QY 671 YLPVVIQIQV-GEPSYQPVSNLTATQOGKVTLKWEAPSAK--KAGSGREVKRIGDGL 726
 Db 648 KETVIKNTINGEPNPYPQVSNLTATQOGKVTLKWDAPSTKTNTATNTARSVDGIRELV 707
 QY 727 FVTIEPAND--VRANEAKVLAADNVAGDNTGYOFLDADHNTFGSVIPA-TGPLFTG-TA 783
 Db 708 LLSVSDAPELLRSGQAEIVLEAHDVNDGSGYIILLDADHDQGVQVIPSHTILMPNCSV 767
 QY 784 SSNLXSANFEYLVPANADPVTTQNIIVTQGEVWIPGGVYDYCIITNPPEASGKMWIAGD 843
 Db 768 PANLF-APPEYTVPENADPSCSPTNIMDGTASVNIAGTYDFAIAPO-ANAKIWIAGQ 825
 QY 844 GGNQPARYDDTFEACKKKTFTMRACMGDGTMEVEDDSPASVYTVVTRDGTIKIEGLT 903
 Db 826 G---PTKDDDYFEAGKKYHFLMKMGSGDGTETITSEGGGSDYTYTVVTRDGTIKIEGLT 882
 QY 904 ATTFEEDGVAAGNHEYCVEVKYTAGVSPKCKDVTVEGNEFAPVQNLGTSSVQKQVTLK 963
 Db 883 ATTFEEDGVATGNHEYCVEVKYTAGVSPKCKDVTVEGNEFAPVQNLGTSSVQKQVTLK 942
 QY 964 WDAPNCTPNPNPNPNPG--TTLSSEFENGIPASWKTIIDADGDGHGKPGNAPGIAGYNS 1022
 Db 943 WDAPNCTPNPNPNPNPG--TTLSSEFENGIPASWKTIIDADGDGHGKPGNAPGIAGYNS 1002
 QY 1023 NGCVYSESGLGIGVLTDPDNYLITPALDIPNGGKLTFWCAODANYASEHYAVYASSTG 1082
 Db 1003 NGCVYSESGLGIGVLTDPDNYLITPALDIPNGGKLTFWCAODANYASEHYAVYASSTG 1062
 QY 1083 NDASNFTNALLEETITAKGVSPKAIIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1142
 Db 1063 NDASNFTNALLEETITAKGVSPKAIIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFY 1122
 QY 1143 IDLDEVEIIXANGKRAIDFTETPESSTHGEAPAWTTTIDADGCGGWCLSSGGLDMLTAHG 1202
 Db 1123 IDLDEVEIIXANGKRAIDFTETPESSTHGEAPAWTTTIDADGCGGWCLSSGGLDMLTAHG 1182
 QY 1203 GSNVYSSFSWNGMALNPDNYLISKDVTKYKYVAVNDGPPGDHYAVI SKTGNAGD 1262
 Db 1183 GTNVVASFNGMALNPDNYLISKDVTKYKYVAVNDGPPGDHYAVI SKTGNAGD 1242
 QY 1263 FTWFEETENGKNGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNSDLNY 1322
 Db 1243 FTWFEETENGKNGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNSDLNY 1302
 QY 1323 ILLDDIQFTMGSPPTDYTVTVYRDGTIKIEGLTETTFEEDGVATGNHEYCVEVKYTAG 1382
 Db 1303 ILLDDIQFTMGSPPTDYTVTVYRDGTIKIEGLTETTFEEDGVATGNHEYCVEVKYTAG 1362

1383 VSPKCVDTVNSTQPNVQNLTAQPNMSDAIKWNAAPSKRAE----- 1428
1363 VSPKCVDTVNSTQPNVQNLTAQPNMSDAIKWNAAPSKRAE----- 1428
1429 -----VLNDFENGIPASWKTDDAGDGNWTTTPPGSSPAGHNSAIC 1473
1421 TIEPANDVRANEAKVLAADNVGNDTGYQFL-LDADHNTFGSVIPATGFTGTASSNL 1479
1474 VSSASHINFEQ--PONPD-----NVLVT--PELSIPGGGTLTFWVCAQDANYASEHYAV 1523
1480 YSA-----NFEYLIPANADPVVTONIIVTGOGEVWIPGG---VYDYCIINPEPAGKXMI 1532
1524 YASSTGNDASNFANALLEVLTAKTIVVTAPEAIRGTRAQGTWYQKTVQVLPAGTKV-APR 1582
1533 -AGDGNQPARVDDTFE-----AGKKTFTWR 1559
1583 HFGCTDFFWINLDDVVITSGNAPSYYTIYRNNTQIASVTTTTRDDPLATGFTYGVK 1642
1560 RAGMGDG-----TDMVEEDSPASYTYTVYRDGTIKIKEGLTETTYRDAGMSAQSHCYCVE 1614
1643 VVYPNGESALETATLNTISLADVTQKPYTLTVVGKTIIVTCQGEAMIVDMNGRBLAAGR 1702
1615 VKYAGVSPKVCVDYIPDGVAUVTAQKPYTLTVVGKTIIVTCQGEAMIVDMNGRBLAAGR 1674
1703 NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1732
1675 NTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 1704

RESULT 13
US-08-570-311-29
Sequence 29, Application US/08570311
Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Progulsk-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guyline
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyronas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570.311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1687 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-29
Query Match 46.2%; Score 4245; DB 2; Length 1687;
Best Local Similarity 51.6%; Pred. No. 6.3e-28;
Matches 920; Conservative 197; Mismatches 443; Indels 224; Gaps 44;
QY 47 FSNFVELTKVETKGTGPASVSIQAFPTGVGVGPEVPAVRKLIAPVPGATPVVRVK-- 104
Db 30 FRMDNLKFTVQTPKGMQAQVPTYTEGVNLSEKGMPTLPILSRSLAVSDTREMKVEVYSSK 89
QY 105 FTEQVYSLNQYGSKLPHQPSMSKSDDEKVPFVYNAAYARKGVGOELTQVEMLTGM 164
Db 90 FIEKNVL-----IAPSKGMIMRNEDPKIPVY-GKSYQNKFPFGELATLDDPEIL 141
QY 165 RGVRIAAALTINPOYDVVANOLKVRNNLEI---EVSFGADEVATQRLYDASPSYPETA 221
Db 142 RDRGQVNVFAPLQINPVTKTRIYTEITVAVSETSEQGRNLLNKKGTGTFAG-----FEDT 196
QY 222 YKQFNDRDVTYDHDGLNTPV-----RMLVAVAGAKPEALKPWLTKAOKGFYLDVHYT 275
Db 197 YKRMF-----MNYEPGRVTPVEBKONGRMIVIAKVEGDIKDFVDWKNQGRLETKVKA 251
QY 276 DEAEVGTINASIKAFIKKY-----NDGLAASAAFPVLAALVGDITVISE--KGKTKTKVT 329
Db 252 EDIASPVTANAIOQFVKQYEKEGND-----LTYVLLVGDHDKDIPAKITPPGIKSDQV- 303
QY 330 DLYYSADVGDYFPEMYTFRMSASSPEELTNIIDKVLAYEKATMDKSVLEKVLIIAGADY 389
Db 304 --YQIVGNHDYNEVFFGRFSCESKEDLKQDITIIHYERNITTEDKWLQALCIIASAG 361
QY 390 SNWSQVQOPTIKY-GMOYYNNQEHGYTDVYNLYKAPYTGCV-----SHLNTGVSF 438
Db 362 GPSADNGESDIOHENVIANLLTQGYTKIIK-----CYDPGVTPKNIIIDAFNGGISL 413
QY 439 ANYTAHGETAWADPLLTTSOLKALTWKDKVFLAIGNCITAQDYQVQPCGEVITRV-- 496
Db 414 VNYTGHGETAWGTSHFGTTHVKQLTNSNQLPFIIDVACVNGDFLFSPMPCFAELMRAQK 473
QY 497 --KEGAYAYIGSSPNSYWGEDIYWSVGANAVFGVQPTFECTSMGSDYDAFTLEDSYNTVN 554
Db 474 DGKPTGVIIASTINGSW-----ASPMRG-----QDEMHEI- 505
QY 555 SIMWAGNLAATHAGNI-----GNTHIGHAYYWEAYH-----VLGDGSVMPYRAM 599
Db 506 -----LCEKHNNIKRTFGVTMGWFAVVEKYKDKGEMKMLDWTVPFGDPSLLVRLTV 558
QY 600 PKTNTYTLPASLPQNASYSIQAS-AGSYVAISKDGLVYGTGVANASGVATVSMTKOITE 658
Db 559 PTQMVTAPAQINLTDSVNVSCDYNAGIATISANGKMFSAVVE-NGTATINLT-GLTN 616
QY 659 NGNYDVVITRSNLYPVIKIQV-GEPSYQPVSNLTATTOGQKVTLKWEAPSAK---KAE 714
Db 617 ESTLTLTVGVNKETVIKTINTNGEPNPYQPVSNLTATTOGQKVTLKWDAPSFKTNATTN 676
QY 715 GSREVKRIGGLFVTTIEPAND-VRAENAKVLAADNVWGDNTGYQFLLDADHNTFGSVIP 773
Db 677 TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWMDGSGYQIILLADHDAQYGVQVP 736
QY 774 A-TGPLEFTG-TASSNLYSANFELYVPANADPVVTONIIVTGOGEVVIPEGVYDICTNP 831
Db 737 SDTHLWPNCSVPANLF-APFEYTVENADPSCPTNMIMDGTASVNIPTAGTYFAIAAP 795
QY 832 EPASGKMVIAGDGNQPARYDDFTFEAGKKYTTTMRAGMGDGTDMVEDEDDSPASYTYT 891

| | |
|------|--|
| 1136 | ANGKRAOFTTFESSTHGEAPAEWITTDADGGQWL--CLSSQQLGWL-----AHG 118 |
| 475 | NCCITAFDY-----VQP-----CFGEVITRVKEGAYAYIGSSPNYSWGEDYTWMSGANAV 526 |
| 1187 | GTNVVASFSGNMALPNPNYLISKDVTCATKVXYVAVNDGPGDH----- 1232 |
| 527 | FGVQTFEGTSGMSYDATFLEDSYNTVN-----S1MWAGNLAAATHAGN 569 |
| 1233 | YAVMISKTGINAGDFTVP--EETPNGKKGARFGLSTEADGAKPQSWIERTVLD,PAGT 1291 |
| 570 | IGNITHIAHY--WEAHYVLGDGSMFYPYAMPKNTNTYLPASLPQQAQSYIQASAGSY 627 |
| 1292 | ---KYVAFRHYNGSDLNYLILDDIQFTMGSGPPTDY----- 1326 |
| 628 | VAISKQGLYCTGVANAS---GVATYSMTKQITENGNYDVITRSNYLPIVKIQIQVG-- 681 |
| 1327 | YTVTRDGTKIKEGLTETTFEEDGVAT-----GNHEYCYEVKYTAGVSPKECVNVT 1376 |
| 682 | EPSYQFVSNLTATTQCKQVTLKWEAPSAKABGSRVKRIGDGLFVTIERANDVRAE 740 |
| 1377 | VDPQFNPVQNLGSAVGQVTLKWDAPNG-----TPNPNPNP 1416 |
| 741 | AKVVLAAADNVGNTGVQPL-LDADHNTFGSVIPATGPLFTG-----TASSNLYSANFE- 793 |
| 1417 | GTUULSESFENGIPASWKTTDADGDGNWTTTPPGTSPAGHNSAICASSAY-INFEG 1475 |
| 794 | -----YLVP-----ANA--DPVV 804 |
| 1476 | PQNPENLYTPELSLPNGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFANALLEVL 1535 |
| 805 | TTQNIIL-----VTG---QOEVIIPGG----- 822 |
| 1536 | TAKTVVTAPEAIRGTRVOGTWYQVQLPAGTKYVAFRHFPGCTDFFWINLDDVEIKANGK 1595 |
| 823 | -----VYDYCIINPEPA-----SGKM-WIADGCGN----- 846 |
| 1596 | RADFTFESSTHGEAPAEWTTDADGGQWLCLSGQLGWLTAHGTTNVVASFSGNMG 1655 |
| 847 | --OPARY-----DDFT--FE----- 857 |
| 1656 | ALNPENLYLSKDVTCATKVXYVAVNDGPGDHVAVMISKTGTNAGDFTVVFETPNGIN 1715 |
| 858 | -----AGKYTMRPAGMGDGTDEVED----- 881 |
| 1716 | KKGARFGLSTEADGAKPQSWIERTVLD,PAGTXYV-APRHYNGSDNLYILDDIQFTMG 1774 |
| 882 | -DSPASYTTVYVRDGTKIKEGLTATTFEEDGVAAGNHEYCYEVKYTAGVSPKCKDVTYE 940 |
| 1775 | SPPTPDYTVYVRDGTKIKEGLTETTFEEDGVATGNHEYCYEVKYTAGVSPKECVNVTVD 1834 |
| 941 | GSNEFAPQNLGSSVGQVTLKWDAPNGTTPNPNPNPNGTTLSESFENGIPASWKTI 1000 |
| 1835 | -PVQFNPVQNLGSAVGQVTLKWDAPNGTTPNPNPNPQT--TTLSESFENGIPASWKTI 1892 |
| 1001 | DADGDGHHKPGNAPG---IAGVNSNGCVVYESFGIGGVLTPDNYLTPALDIPNGCK 1057 |
| 1893 | DADGDGNWTTTPPGGTSPAGHNSAICVSSASY-INFEQPQNPENLYTPELSLPGGT 1951 |
| 1058 | LTFWCAQDANYASEHYAVVASSTGNDAGNFTNALLEETITAKG-VRSKPAIRG-RIQGT 1115 |
| 1952 | LTFWCAQDANYASEHYAVVASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGT 2011 |
| 1116 | WRQKTVLD,PAGTKYVAFRHFQSDMEVIDLDEVEIKANGKRADFTTFESSTHGEAPASW 1175 |
| 2012 | WYQKTVQLPAGTKYVAFRHFPGCTDFWINLDEVEIKANGKRADFTTFESSTHGEAPASW 2071 |
| 1176 | TTDADGGQWLCLSSGQLDMLTAHGGSNVSSFSWNGMALNPONLYLSKDVGTATKYK 1235 |
| 2072 | TTDADGGQWLCLSSGQLDMLTAHGGTNVVASFSWNGMALNPONLYLSKDVGTATKYK 2131 |
| 1236 | XYAVVNDGPGDHVAVMISKTGTNAGDFTVVFETPNGKKGARFGLSTEANGAKPQSV 1295 |
| 2132 | XYAVVNDGPGDHVAVMISKTGTNAGDFTVVFETPNGKKGARFGLSTEANGAKPQSV 2191 |

RESULT 15

US-09-482-500A-1

; Sequence 1, Application US/09482500A

; Patent No. 6627193

GENERAL INFORMATION:

; APPLICANT: Travis, James

APPLICANT: Imamura, Takahisa

APPLICANT: Potempa, Jan

1. TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION

FILE REFERENCE: 235.00160101

1 ; CURRENT APPLICATION NUMBER: US/09/482,500A
2 ; CURRENT FILING DATE: 2001 12 17

: CURRENT FILING DATE: 2001-12-17
 : PRIOR APPLICATION NUMBER: US 60/

; PRIOR FILING DATE: 1999-01-13
 ; PRIOR APPLICATION NUMBER: US 60/115,863

: NUMBER OF SEQ ID NOS: 4
: ERROR FILING DATE: 1993-01-13

NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 1477

```

; TYPE: PRT

```

ORGANISM: *Porphyromonas gingivalis*

US-09-482-500A-1

Query Match

| | | |
|-----------------------|-------|---------------------|
| Best Local Similarity | 55.3% | Pred. No. 1.7e-275; |
| | | 50000 100000; |

| | | | | | | | | | |
|---------|------|--------------|------|------------|------|--------|------|------|-----|
| Matches | 876; | Conservative | 156; | Mismatches | 353; | Indels | 199; | Gaps | 38; |
|---------|------|--------------|------|------------|------|--------|------|------|-----|

QY 240 TPV-----RMLVVAGAKFKEALKPWL TWKAOKGEYLDVHYTDEAEVGTTNASTKAEFTHK 293

[illegible]

Z Z FZKQAGSIOIA VYVRAUCISODINDF VDWNKNQRGDKIEVKVAEDIASPVIANAIQQFVRQ 6I

QY 294 KY----NDGLAASAApVFLALVGDTDVISGE--KGKKTKKVTDTI.VYSAVNDGNYEPFEMVTE 347

```
db 62 EYEKEGND-----I.TYV.I.VGNHKDTPAKITRGTKSDOV--YCGTYCNDHVAHEITIC 111
```

02 2123CAND
E11VDEFGHNDIFARTIFGIASDQV---IGQIVGNDHYNEVFIG 1111

QY 348 KMSASSFEELINIDKVLMIYEKATNPDKSYLEKVLIIAGADYSWNSQVGQPTIKY-GMQY 406

Db 172 NLLQYGYTKIIR-----CYDPGVTPKMIIDAFNGGISLVNVTGHGSETANGTSHFG 223
QY 457 TSOLKALTNKDKYFLAIGNCICITAQFYQVQPCFEVITRV-----KEGAYAYIGSSNSY 512
Db 224 TTHVKQLTNSQOLFIEDVACVNDGDFLFEMPCFBAELMRAQOKGKPTGTVAIIASTINQS 283
QY 513 WGEDYYWSVGANAVFGVQPTFEGTSMGSDYATFLEDSYNTVNSIMWAGNLAATHAGNI-- 570
Db 284 W-----ASPMRG-----QDEWNEI-----LCEKHPNNIKR 308
QY 571 --GNITHGAHYWEAH-----VLGDOSVMPYRAMPKNTYTYTPASLPQNOAS 617
Db 309 TFGGVTVWNGMFMAMEYKXKQGEKMLDWTWFGDPSLVLTVLPTKMQVTAQAQINLTDAS 368
QY 618 YSIQAS--AGSYVALSKDGLVYGTGVANASGVATVSMTKQITENGNYDVVITRSNYLPIVK 676
Db 369 VNVSCDYNGAIAATISANGKEGSAVBE--NGTATINLT--GLTNESTLTLTVVGYNKETVIK 426
QY 677 QIQV--GEPSYQPVNSLNTATTQOKVTLKWEAPSAK---KAEGSREVKRIQDGLFVTIEP 732
Db 427 TINTNGEPNYPQVNSLNTATTQOKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSD 486
QY 733 AND--VRANEAKVVLADNVWGDNTGYQFLDADHNTFGSVIPA--TGPLFTG--TASSNIYS 789
Db 487 APELLRSQQAIEI VLEAHDVWWDGSGYQILLDADHDQYGVIPSDTHTLWPNCSVPANLF-- 545
QY 790 ANFEYLVPANADPVVITQNIIVTQGEWJIPGGVYDYCIITNPEPASGKMMIADGGNQPA 849
Db 546 ARFEYTVPENADPSCSTNNIMDGTSVNI PAGTYDFAIAAPO--ANAKIWIAGQG---PT 601
QY 850 RYDDFTFEAGKKYTFMRRRAGMGDGTMEVEDDSPASYTYTVYRDGTIKIKEGLTATTFEE 909
Db 602 KEDDYVFEAGKXHYFLMKMGSGDGTETISEGGSDDYTVTVYRDGTIKIKEGLTATTFEE 661
QY 910 DGVAAGNHEXCVKVKYTAGVSPKCKDVTVEGNEFAPVONLTGSSVGVQKVTILKWDAPNG 969
Db 662 DG VATGNHEXCVKVKYTAGVSPKCKDVTVEGNEFAPVONLTGSAVGVQKVTILKWDAPNG 721
QY 970 TENPNPNPNPG--TTLISESFENGIPASWKTIIDADGDGHGKPGNAPGIAGYNSNGCVYS 1028
Db 722 TENPNPNPNPNPCTTLISESFENGIPASWKTIIDADGDGHGKPGNAPGIAGYNSNGCVYS 781
QY 1029 ESFGLGGIGVLPDNYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1088
Db 782 ESFGLGGIGVLPDNYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841
QY 1089 TWALLEETITAKGVRSPKAIRGIQGTWRQKTVDLPAKTYVAFRHFQSTDMFYIDLDEV 1148
Db 842 TWALLEETITAKGVRSPKAIRGIQGTWRQKTVDLPAKTYVAFRHFQSTDMFYIDLDEV 901
QY 1149 EIKANGKRADFTTFESSSTHGEAPAEWTTIDADGDGQGMCLSSGQDMLTAHGGSNVYS 1208
Db 902 EIKANGKRADFTTFESSSTHGEAPAEWTTIDADGDGQGMCLSSGQDMLTAHGGSNVVA 961
QY 1209 SFSWNGMALNPONYLISKDVTGATKYKYIYAVNDGPPGDHYAVMISKTGTNAGDFTVWFE 1268
Db 962 SFSWNGMALNPONYLISKDVTGATKYKYIYAVNDGPPGDHYAVMISKTGTNAGDFTVWFE 1021
QY 1269 ETPNGINKGARFGLSTEANGAKPQSVLERTVDLPAGTKYVAFRHYNCSDNYILLDDI 1328
Db 1022 ETPNGINKGARFGLSTEANGAKPQSVLERTVDLPAGTKYVAFRHYNCSDNYILLDDI 1081
QY 1329 QFTMGSSPTPTDYTYTVYRDGTIKIKEGLTETTFEDGVATGNHEXCVKVTAGVSPKPC 1388
Db 1082 QFTMGSSPTPTDYTYTVYRDGTIKIKEGLTETTFEDGVATGNHEXCVKVTAGVSPKPC 1141
QY 1389 VDVTVNSTQFNPVQNIATAEQAPNSMDAILKNAPASKRAB----- 1428
Db 1142 VNVVTNPQTQFNPVQNLKAO--PDGGDVLLKWEAPSAKKTGSGREVKRIQDGLFVTIEPAN 1199
QY 1429 -----VLNEDFENGIPASWKTIIDADGDNWNTTTPPGSSSFAGHNSAICVSSASH 1479

Db 1200 DVRAEAKVVLADNVWGDNTGYQFL--LDADHNTFGSVIPATGFLFTGTASSNLYSA--- 1255
QY 1480 INPEG--PQNPDP-----NVLVT--PELSLPGGSTLTFWVCAQDANYASEHYAVYASSTG 1529
Db 1256 -NFEYLLIPANADPVVITQNIIVTQGEWVIFPG--VYDYCIITNPEPASGKMMI--AGDGG 1310
QY 1530 NDASNANFANALLEEVLTAKTAVVTAPEALRGTRAQCTWYQKTVQLPAGTKYV--APRHFQCTD 1588
Db 1311 NQPARYDDFTFE-----AGKKYTFMRRRAGMGD 1338
QY 1589 FFWINLDDVVITSGNAPSYYTYTYIRNNTQIASGVTTETTYRDPDLATGFYTYGVKVVYPNG 1648
Db 1339 G-----TDMVEEDDSPASYTYTVYRDGTIKIKEGLTETTYRDAGMSAQSHCYEVKTYAAG 1393
QY 1649 ESALETATLNTISLADVTAGKPYTLTVGKTIITVTCQGEAMIYDMNGRRRLAAGRNTVVYT 1708
Db 1394 VSPKVCVDYIPDGVDVDTAGKPYTLTVGKTIITVTCQGEAMIYDMNGRRRLAAGRNTVVYT 1453
QY 1709 AOGGHYAVMVVVDGKSYVEKLAVK 1732
Db 1454 AOGGYAVMVVVDGKSYVEKLAVK 1477

Search completed: May 18, 2004, 11:36:55
Job time : 36.7501 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:32:49 ; Search time 69.1557 Seconds
(without alignments)
6969.043 Million cell updates/sec

Title: US-08-353-485-10

Perfect score: 9179

Sequence: 1 MRKLLLLIAASLLGVGLYLAQ.....HYAVMVVDGKSYVEKLAVK 1732

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 9167 | 99.9 | 1732 | 14 | US-10-229-066-11 |
| 2 | 4272 | 46.5 | 1706 | 14 | US-10-229-066-11 |
| 3 | 2702 | 29.4 | 509 | 15 | US-10-387-977-101 |
| 4 | 2254 | 24.6 | 419 | 15 | US-10-174-695-5 |
| 5 | 1776 | 19.3 | 419 | 15 | US-10-174-695-3 |
| 6 | 1241 | 13.5 | 231 | 15 | US-10-174-695-6 |
| 7 | 541 | 5.9 | 736 | 15 | US-10-387-977-105 |
| 8 | 527.5 | 5.7 | 491 | 15 | US-10-387-977-100 |
| 9 | 521.5 | 5.7 | 136 | 15 | US-10-174-695-4 |
| 10 | 410.5 | 4.5 | 507 | 15 | US-10-387-977-99 |
| 11 | 309 | 3.4 | 1728 | 12 | US-10-282-122A-56997 |
| 12 | 303.5 | 3.3 | 2435 | 12 | US-10-282-122A-47453 |
| 13 | 300.5 | 3.3 | 2468 | 12 | US-10-282-122A-66335 |
| 14 | 300.5 | 3.3 | 2468 | 12 | US-10-246-330-4 |
| 15 | 293 | 3.2 | 1638 | 12 | US-10-206-576-258 |

| | | | | | | |
|----|-------|-----|-------|----|----------------------|-------------------|
| 16 | 293 | 3.2 | 1638 | 12 | US-10-206-576-262 | Sequence 262, App |
| 17 | 293 | 3.2 | 1638 | 12 | US-10-206-576-266 | Sequence 266, App |
| 18 | 278.5 | 3.0 | 6310 | 12 | US-10-282-122A-67793 | Sequence 67793, A |
| 19 | 264 | 2.9 | 1283 | 15 | US-10-369-493-22616 | Sequence 22616, A |
| 20 | 257 | 2.8 | 4327 | 15 | US-10-369-493-10178 | Sequence 10178, A |
| 21 | 247.5 | 2.7 | 1325 | 12 | US-10-282-122A-44953 | Sequence 44953, A |
| 22 | 247 | 2.7 | 3930 | 12 | US-10-282-122A-46817 | Sequence 46817, A |
| 23 | 239 | 2.6 | 1741 | 9 | US-09-971-536-68 | Sequence 68, Appl |
| 24 | 238 | 2.6 | 1649 | 15 | US-10-369-493-18460 | Sequence 18460, A |
| 25 | 237.5 | 2.6 | 3217 | 9 | US-09-797-862-33 | Sequence 33, Appl |
| 26 | 237.5 | 2.6 | 3217 | 12 | US-10-282-122A-61210 | Sequence 61210, A |
| 27 | 232.5 | 2.5 | 1751 | 12 | US-10-282-122A-53633 | Sequence 53633, A |
| 28 | 231 | 2.5 | 1308 | 12 | US-10-282-122A-44566 | Sequence 44566, A |
| 29 | 229.5 | 2.5 | 1483 | 12 | US-10-282-122A-51483 | Sequence 51483, A |
| 30 | 227 | 2.5 | 3705 | 12 | US-10-282-122A-77944 | Sequence 77944, A |
| 31 | 226.5 | 2.5 | 2457 | 12 | US-10-282-122A-49854 | Sequence 49854, A |
| 32 | 226 | 2.5 | 5559 | 12 | US-10-282-122A-75177 | Sequence 75177, A |
| 33 | 225.5 | 2.5 | 1333 | 12 | US-10-282-122A-55546 | Sequence 55546, A |
| 34 | 225 | 2.5 | 26926 | 9 | US-09-759-508B-2 | Sequence 2, Appl |
| 35 | 224 | 2.4 | 1521 | 14 | US-10-238-075-909 | Sequence 909, App |
| 36 | 223.5 | 2.4 | 1920 | 12 | US-10-282-122A-71413 | Sequence 71413, A |
| 37 | 223 | 2.4 | 1541 | 10 | US-09-839-996-3 | Sequence 3, Appl |
| 38 | 223 | 2.4 | 1541 | 12 | US-10-645-655-3 | Sequence 3, Appl |
| 39 | 223 | 2.4 | 1541 | 14 | US-10-080-505-3 | Sequence 3, Appl |
| 40 | 223 | 2.4 | 4283 | 15 | US-10-080-334-275 | Sequence 275, App |
| 41 | 222.5 | 2.4 | 1289 | 12 | US-10-282-122A-44901 | Sequence 44901, A |
| 42 | 221.5 | 2.4 | 2234 | 12 | US-10-282-122A-46565 | Sequence 46565, A |
| 43 | 220 | 2.4 | 1539 | 14 | US-10-085-959-253 | Sequence 253, App |
| 44 | 219.5 | 2.4 | 1953 | 12 | US-10-282-122A-43730 | Sequence 43730, A |
| 45 | 219 | 2.4 | 2902 | 12 | US-10-282-122A-59042 | Sequence 59042, A |

ALIGNMENTS

RESULT 1

US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric S.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/006673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match 99.9%; Score 9167; DB 14; Length 1732;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1729; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLLLLIAASLLGVGLYLAQSAKIKLIDAPTTRTTCTNNKFKQFDASFSFNEVELTKVETK 60
Db 1 MRKLLLLIAASLLGVGLYLAQSAKIKLIDAPTTRTTCTNNKFKQFDASFSFNEVELTKVETK 60
QY 61 GGTFFASVSTPGAFPTGCVGSPEVPAVRKLIAPVAGATPVVRKSFTEQVYSLNQVSEKL 120
Db 61 GGTFFASVSTPGAFPTGCVGSPEVPAVRKLIAPVAGATPVVRKSFTEQVYSLNQVSEKL 120

121 MPHQSMSKSDPDKPVFPVNAAYARKFGVQBELTQVEMLGTMRGVRIAAALITNPQYD 180
121 MPHQSMSKSDPDKPVFPVNAAYARKFGVQBELTQVEMLGTMRGVRIAAALITNPQYD 180
181 VVANQLKVRNNIEIEVSFQGADEVATORLYDASFPFETAYKOLFNRDVVTDHGDLYNT 240
181 VVANQLKVRNNIEIEVSFQGADEVATORLYDASFPFETAYKOLFNRDVVTDHGDLYNT 240
241 PVRMLVAGAKFKBALPWLTKAOKGFLVDVHTDEAEVGTNNASLKAFTHKKYNDGLA 300
241 PVRMLVAGAKFKBALPWLTKAOKGFLVDVHTDEAEVGTNNASLKAFTHKKYNDGLA 300
301 ASAAPVFLALVGDTRVDSGEKGTKKVTDLYYSAVDGDPPEMYTFRMSASSPEELTNI 360
301 ASAAPVFLALVGDTRVDSGEKGTKKVTDLYYSAVDGDPPEMYTFRMSASSPEELTNI 360
361 IDKVLMEKATMPDKSYLEKULLTAGADYSNWSQVGTIKYGMQYNNQSHGVTDVNY 420
361 IDKVLMEKATMPDKSYLEKULLTAGADYSNWSQVGTIKYGMQYNNQSHGVTDVNY 420
421 LKAPYTCYSHLNTGVSFANYTAGHSETAMADPLTTLSQKALTNKDKYFLAIGNCCITA 480
421 LKAPYTCYSHLNTGVSFANYTAGHSETAMADPLTTLSQKALTNKDKYFLAIGNCCITA 480
481 QFDYVQPCFGEVITRVKEKGAAYIGSPNSYWGEDYYSVGANAFVGVQPTPEGTSMGS 540
481 QFDYVQPCFGEVITRVKEKGAAYIGSPNSYWGEDYYSVGANAFVGVQPTPEGTSMGS 540
541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSMVPYRAMP 600
541 YDATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYWEAYHVLGDGSMVPYRAMP 600
601 KTWNTYTLPASLPQNOASYSIOASAGSVVAISKDGLVGTGVANASGVATVSMTKQITENG 660
601 KTWNTYTLPASLPQNOASYSIOASAGSVVAISKDGLVGTGVANASGVATVSMTKQITENG 660
661 NYDVVITRNSYLPVVKIQI QVGEPSYQVNSLNTATTQKQVTLKWEAPSAKAGSREV 720
661 NYDVVITRNSYLPVVKIQI QVGEPSYQVNSLNTATTQKQVTLKWEAPSAKAGSREV 720
721 RIGDGLFVTLEPANDVRANEAKVLAADNVGMDNTGYOFLDADHNTFGSVIPATGLPFT 780
721 RIGDGLFVTLEPANDVRANEAKVLAADNVGMDNTGYOFLDADHNTFGSVIPATGLPFT 780
781 GTASSNLSANFEYLIPANADPVVTTQNIIVTGQEVVPGVVDYCIITNPEPASGKWI 840
781 GTASSNLSANFEYLIPANADPVVTTQNIIVTGQEVVPGVVDYCIITNPEPASGKWI 840
841 AGDGNQPARYDDFTFAGKKTFTMRAGMGDDGTDMEVEDDSDPASVYTYVYRDGTXKE 900
841 AGDGNQPARYDDFTFAGKKTFTMRAGMGDDGTDMEVEDDSDPASVYTYVYRDGTXKE 900
901 GLTATTPEEDGVAAGNHEYCVVEKYTAGVSPKVCCKVTVSGNEFAFVQNLTGSSVQCKV 960
901 GLTATTPEEDGVAAGNHEYCVVEKYTAGVSPKVCCKVTVSGNEFAFVQNLTGSSVQCKV 960
961 TLKWDAPNGTNPNNPNNPNTLSSFPENGIPASWKTIDADGDGHGKPGNAPGIAGY 1020
961 TLKWDAPNGTNPNNPNNPNTLSSFPENGIPASWKTIDADGDGHGKPGNAPGIAGY 1020
1021 NSNGCVVSEFGLGGIGVLPDNYLITPDLDPNGGKLTFWVCADQANYASEHYAVYASS 1080
1021 NSNGCVVSEFGLGGIGVLPDNYLITPDLDPNGGKLTFWVCADQANYASEHYAVYASS 1080
1081 TGNDAASFTNALLLEETITAKGVRSFKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
1081 TGNDAASFTNALLLEETITAKGVRSFKAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDM 1140
1141 FYIDLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQGGQWLCSSGGQDLWLT 1200
1141 FYIDLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQGGQWLCSSGGQDLWLT 1200
1201 HGSNNVSSFSWNGMALNPNDYLLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGNA 1260

1201 HGSNNVSSFSWNGMALNPNDYLLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGNA 1260
1261 GDFTVFEEETPENGKNGGARFGLSTEANGAKPQSVWIERVTDLPAGTKYVAFRHYNGSDL 1320
1261 GDFTVFEEETPENGKNGGARFGLSTEANGAKPQSVWIERVTDLPAGTKYVAFRHYNGSDL 1320
1321 NYILLDDIQTFTMGSGPTPTDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHEYCVVKYT 1380
1321 NYILLDDIQTFTMGSGPTPTDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHEYCVVKYT 1380
1381 AGVSPKCVDVTVNSTQFNPNQNLTAEOAPNSMDAILKMNAPASKRAEVLNEDFENGIPA 1440
1381 AGVSPKCVDVTVNSTQFNPNQNLTAEOAPNSMDAILKMNAPASKRAEVLNEDFENGIPA 1440
1441 SWKTIDADGNGNWTTPPPGGSSFAGHNSAICVSSASHINFEQPQNDNYLVTPELSLP 1500
1441 SWKTIDADGNGNWTTPPPGGSSFAGHNSAICVSSASHINFEQPQNDNYLVTPELSLP 1500
1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEARGTR 1560
1501 GGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEARGTR 1560
1561 AOGTWTQKTVQLPAGTKYVAFRHFQCTDFFWINLDDVWITSGNAPSXYTYIYRNTQIAS 1620
1561 AOGTWTQKTVQLPAGTKYVAFRHFQCTDFFWINLDDVWITSGNAPSXYTYIYRNTQIAS 1620
1621 GVTETTYRDDPLATGFTYGVKVVYNGESAIETATLNTSLADVTAQKPYTLTVVGKTI 1680
1621 GVTETTYRDDPLATGFTYGVKVVYNGESAIETATLNTSLADVTAQKPYTLTVVGKTI 1680
1681 TVTCQGEAMLYDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK 1732
1681 TVTCQGEAMLYDMNGRRLAAGRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK 1732

RESULT 2
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRN
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 46.5%; Score 4272; DB 14; Length 1706;
Best Local Similarity 50.9%; Pred. No. 0;
Matches 932; Conservative 204; Mismatches 456; Indels 240; Gaps 47;
QY 6 LLTAASLLGVGLYQAQSAKIKLDAPTTRTCTTNSFKQFDASFSNEVELTKVETKGGTFA 65
10 IALCSSILGGWAFQAQQTGLGRNPVLLSTQGSVTK--VQFRMDNLKFTETVQTPKGI-- 65
66 SVSIPGAPT-----GEVSGPEVPAVKLIAVPVGAATPVVRVKS--PTEQVYSLNQYV 116
66 -----QGVFTYTBGVNLSEKGMPTLPIILSRSLAVSDTRMKVEVSVSSKTEKKNVL---- 116

117 SEKIMPHQPSMSKDDPEKVPFVYNAAAYARKFVGOELTQVEMLGTMRGVRIAALFINP 176
117 ---IAPSKGIMRNEDPKPIPVYV-KTYSQNKFFPGCEIATLDDPPFLTRDVRGOVNFAP 172
177 VOYDVVANQIKVRNIEI---EVSFGADEVATQRLYDASPSYFETAYKQLFNDRDYTD 233
173 LOYNPVTKLRIIYETIIVAVSETSEQKNTLNKKGTFAG-----FETYKRMF-----MN 222
234 HGLDYNTPV-----RMLVVAGAKFKEALPWLTKAOKGYLDVHVHTDEAVGTINASI 287
223 YEGRYTPVEEKQKQRMIVIAKKYBGDIKDFVDMKNQGRLEKVEKVAEDIAFVPTANAI 282
288 KAFIHKY-----NDGLAASAPFLALVGDTVISGE--KGKTKKVTLDLYYSAVGDYF 341
283 QQFVKQYEKEGND-----LTVLLIGDHKOIPAKITPGIKSDQV---YGVQVGNHXY 332
342 PEMYTRMSASSPELTNIIIDKVLMEKATMPDKSYLEKVLILLIAGADYSWNSQVQPTIK 401
333 NEVFIQRFSCSKEDLKTQIDRTIHERNITTEDKMLGQALCIASAEKGGPSADNGESDIQ 392
402 Y-GMOYVYNOEHGVTVDVYNLKPATYTCY-----SHLNTGVSFANYTAHGSSETAW 450
393 HENVIANLLTOYGYTKLIK-----CYDPGVPKNIIDAFNGGISLANITGHGSSETAW 444
451 ADPLLTSSQLKLTNKDKYFLAIGNCCITAQFDYVQPCFGEVITRV-----KEKGAVAYIG 506
445 GTSHFGTTHVKQLTNSQNLFPFIDVACVNGDFLFSMPCFAEALMRAQCKGKPTGTVALIA 504
507 SSPNSWGEDYVSVGANAVFGQPTFEGTMSGSDATFLEDYNNVNSIMWAGNLAATH 566
505 STINQSW-----ASPMRG-----QDEMNEI-----LCEKH 529
567 AGNI-----GNITHI GAHYWEAYH-----VLGDGSMVPYRAMPKNTYTLPASL 611
530 PNNIKTFGVTWNGFMFAMVEKKYKKGEMLDTWTFGDPSSLVRLVLPTKMQVTPAQI 589
612 PQOASYSIQAS--AGSYVAISKDGLVYGTGVANASGVATVSMTKQITENGYNDVWITRSN 670
590 NLTDASVNSCDYNGAIAIATISANGKMFSGSAVVE-NGTATINLT-GLTNESTLTLTVVGN 647
671 YLPVTKIQIOV-GEPSYQOPVSNLTATQCKVTLKWEAPSAK---KABGSREVRITGDGL 726
648 KETVIKINTNGEPNYPQVSNLTATQCKVTLKWDAPSTKTNATNTARSVDGIRELV 707
727 FVTEPAND-VRAEAKVLAADNVMDNTGYOFLDADHNTFGSVIPA-TGPLETG-ITA 783
708 LLSVSDAPELLRSQAEIVLEAHDVWNGSGYQILLDADHDQGVIPSDTHLWENCNV 767
784 SSNLSANPEYLVPANADPVVTTQNTIIVTQGEVVIIPGGVYDYCIITNPEPASCKWMIAGD 843
768 PANLF-APPEYTVPENADPSCPTNMIMDGTASVNIIPAGTYDFAIAAPO-ANAKIWIAGQ 825
844 GGNQPARYDDTFEAGKKYFTFMRAGMGDGTMEVEDDSPASYTYTVYRDGTGKKEGLT 903
826 G---PTKEDDYVFEAGKKYHFLMKMGSGDGLTITISEGGSDTYTVYRDGTGKKEGLT 982
904 ATTFEEDGVAAGNHEYCVBEVKYTAGVSPKCVKDVTVEGSNEFAPVONLTGSSVGOKVTLK 963
883 ATTFEEDGATGNHEYCVBEVKYTAGVSPKCVKDVTVEGSNEFAPVONLTGSAVGQVTLK 942
964 WDAENGTNPNNPNPNPG---TTLSSEFENGIPASWKTIDADGCGHGWKPGNAPGIAGY 1020
943 WDAENGTNPNNPNPNPNPGTTLSSEFENGIPASWKTIDADGCGHGWKPGNAPGIAGY 1002
1021 NSNGCVYSEFGLGGIGVLTDPNLYLTALDLPNGCKLTFWCAQADANYASHYAVTASS 1080
1003 NSNGCVYSEFGLGGIGVLTDPNLYLTALDLPNGCKLTFWCAQADANYASHYAVTASS 1062
1081 TGNDSNFTNALLEETITAKGVRSKPAIRGRIGQWTRQKTVDLPAGTKYVAFRHFQSTDM 1140
1063 TGNDSNFTNALLEETITAKGVRSKPAIRGRIGQWTRQKTVDLPAGTKYVAFRHFQSTDM 1122

1141 FYIDLDEVEIKANGKRADFTETESSTHGEAPAEWTTIDADGQGWCLSSGOLDWLTA 1200
1123 FYIDLDEVEIKANGKRADFTETESSTHGEAPAEWTTIDADGQGWCLSSGOLDWLTA 1182
1201 HGSNNVVSFSWNGMALNPDNYLISKDVTGATKVYKYAYVNDGPPGDHYAVMISKGTGNA 1260
1183 HGTNTVVSFSWNGMALNPDNYLISKDVTGATKVYKYAYVNDGPPGDHYAVMISKGTGNA 1242
1261 GDTVVVPEETPENGINKGARGFLSTANGAKPOSWMIERTVDLPAGTKYVAFRHYNCSDL 1320
1243 GDTVVVPEETPENGINKGARGFLSTANGAKPOSWMIERTVDLPAGTKYVAFRHYNCSDL 1302
1321 NYLLDDIOFTMGSGSPPTDYTVYVYRDGTGKKEGLTETTFEEDGVATGNHEYCVBEVKYT 1380
1303 NYLLDDIOFTMGSGSPPTDYTVYVYRDGTGKKEGLTETTFEEDGVATGNHEYCVBEVKYT 1362
1381 AGVSPKCVDVTVNVSQTFNPVQNLTAQAPNSMDAILKNAPASKRAE-----1428
1363 AGVSPKCVNVTNVSQTFNPVKNLKAQ--PDGDDVVLKWEAPSAKKTGESSREVKRIGDGL 1420
1429 -----VLNEDFENGIPASWKTIDADGDNWNTTTPPGGSSPAGHNSA 1471
1421 FVTEPANDVRAEAKVLAADNVMDNTGYOFL-LDADHNTFGSVIPATGPIFTGTASS 1479
1472 ICVSSASHINFEG--PONPD-----NYLVT--PELSLPGGGTLTFWVCAQADANYASEHY 1521
1480 DLYSA-----NFESLIPANADPVVTTQNTIIVTQGEVVIIPGG---VVDYCIITNPEPASGKM 1532
1522 AVYASSTGNDASNFANALLEEVLTAKTVVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYV-A 1580
1533 WI-AGDCGNQPARYDDTFE-----AGKXYTFT 1559
1581 FRHFGCTDFWNLDDVWITSGNAPSYYTYIYRNNTQIASGTETTYTRPDLATGTYTYG 1640
1560 MRRAGMDG-----TMEVEDDSPASYTYTVYRDGTGKKEGLTETTYTRDAGMSAQSHYEC 1614
1641 VKVYVNGESAIETATLNTISLADVTAPKPYTLTVVGKTTITVTCQEAMLYDMNGRLAA 1700
1615 VEVKYTAGVSPKCVDYIIPGVADVTAQKPYTLTVVGKTTITVTCQEAMLYDMNGRLAA 1674
1701 GRNTVVYTAQGHYAVMVVVDGKSYVEKLAVK 1732
1675 GRNTVVYTAQGGYVAVMVVVDGKSYVEKLAIK 1706
RESULT 3
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

```
Query Match      29.4%; Score 2702; DB 15; Length 509;
Best Local Similarity 100.0%; Pred. No. 6.3e-199;
Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 DVTYDHGDLNTPVPMIYVAGAKFKEALKPMLTWKAQKGFYLDVHYTDEAEVGTINASIK 288
DB 1 DVTYDHGDLNTPVPMIYVAGAKFKEALKPMLTWKAQKGFYLDVHYTDEAEVGTINASIK 60

QY 289 AFHKYNDGLAASAPVFLALVGDVTSGBKGKTKKVTDLIYSYAVDGDYFPEMYTFR 348
DB 61 AFHKYNDGLAASAPVFLALVGDVTSGBKGKTKKVTDLIYSYAVDGDYFPEMYTFR 120

QY 349 MSASSPEELTNIIDKVLMEYKATMPDKSKYLEKVLIIAGADYSWNSQVQPTIKYGMQYYY 408
DB 121 MSASSPEELTNIIDKVLMEYKATMPDKSKYLEKVLIIAGADYSWNSQVQPTIKYGMQYYY 180

QY 409 NOEHGYTDVYNYLKAPYTCGYSHLNTGVSFANYTAHGETAWADPLLITTSQIKALTNKDK 468
DB 191 NOEHGYTDVYNYLKAPYTCGYSHLNTGVSFANYTAHGETAWADPLLITTSQIKALTNKDK 240

QY 469 YFLAIGNCCITTAQFDYVQPCFGEVITRVEKEGAYIIGSSPNSYWGEDYYWSVGANAVFG 528
DB 241 YFLAIGNCCITTAQFDYVQPCFGEVITRVEKEGAYIIGSSPNSYWGEDYYWSVGANAVFG 300

QY 529 VQPTFEGTSMGSDYATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVL 588
DB 301 VQPTFEGTSMGSDYATFLEDSYNTVNSIMWAGNLAATHAGNIGNITHIGAHYYWEAYHVL 360

QY 589 GDGSMVPYRAMPKNTYTLPSLPONQASYSIQASAGSYVAISKDGLVLYGTGVANASGVA 648
DB 361 GDGSMVPYRAMPKNTYTLPSLPONQASYSIQASAGSYVAISKDGLVLYGTGVANASGVA 420

QY 649 TVSMTKQITENGNDVVTIRSNLFPVIKQIVGSPSPYQPVSNLTATTQGGKVLTKWEAP 708
DB 421 TVSMTKQITENGNDVVTIRSNLFPVIKQIVGSPSPYQPVSNLTATTQGGKVLTKWEAP 480

QY 709 SAKAEGSREVKRIGDGLFVTIEPANDVR 737
DB 481 SAKAEGSREVKRIGDGLFVTIEPANDVR 509

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; PRIOR FILING DATE: 2002-06-18
; PRIOR FILING DATE: 2000-12-21
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match      24.6%; Score 2254; DB 15; Length 419;
Best Local Similarity 99.8%; Pred. No. 1.4e-164;
Matches 418; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 738 ANEAKVVLAAADNVGNDTGYQFLLDADHNTFGSVIPATGTLTASSNLYSANFEYLVP 797
DB 1 ANEAKVVLAAADNVGNDTGYQFLLDADHNTFGSVIPATGTLTASSNLYSANFEYLIP 60
```

```
QY 798 ANADPVVVTQNIIVTGOGEVVI PGVYDYCTITNPEPAGKMWIAGDGNQPARVDDTFFE 857
DB 61 ANADPVVVTQNIIVTGOGEVVI PGVYDYCTITNPEPAGKMWIAGDGNQPARVDDTFFE 120

QY 858 AGKYTFTMRRAGMGDGTMEVEDDSDPASVYTYTVYRDGTKIKEGLTATTTFEEDGVAAGNH 917
DB 121 AGKYTFTMRRAGMGDGTMEVEDDSDPASVYTYTVYRDGTKIKEGLTATTTFEEDGVAAGNH 180

QY 918 EYCEVYKYTAGVSPKVKDVTVEGSNEFAPVQNLTGSSVGQKVTWKWDAPNGTNPNP 977
DB 181 EYCEVYKYTAGVSPKVKDVTVEGSNEFAPVQNLTGSSVGQKVTWKWDAPNGTNPNP 240

QY 978 NPNPGLTTLSSFFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSGCVYSESGLGGIG 1037
DB 241 NPNPGLTTLSSFFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSGCVYSESGLGGIG 300

QY 1038 VLTPDNLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNTWALLBETI 1097
DB 301 VLTPDNLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNTWALLBETI 360

QY 1098 TAKVRSFKAIRGRIQGTWRQKTVDDLPAQTKYVAFRHFQSTDMFYIDLDEVEIKANGKR 1156
DB 361 TAKVRSFKAIRGRIQGTWRQKTVDDLPAQTKYVAFRHFQSTDMFYIDLDEVEIKANGKR 419

RESULT 5
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; PRIOR FILING DATE: 2002-06-18
; PRIOR FILING DATE: 2000-12-21
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match      19.3%; Score 1776; DB 15; Length 419;
Best Local Similarity 79.2%; Pred. No. 9.4e-128;
Matches 336; Conservative 31; Mismatches 47; Indels 10; Gaps 6;

QY 738 ANEAKVVLAAADNVGNDTGYQFLLDADHNTFGSVIPATGTLTASSNLYSANFEYL 795
DB 1 SGQAEIVLEAHDVWMDGSGYQILLDADHDQYGVIPSDTHTLWPNCSVPANLF-APFEYT 59

QY 796 VPANADPVVVTQNIIVTGOGEVVI PGVYDYCTITNPEPAGKMWIAGDGNQPARVDDFT 855
DB 60 VPANADPVVVTQNIIVTGOGEVVI PGVYDYCTITNPEPAGKMWIAGDGNQPARVDDFT 115

QY 856 FEAGKKYTFTMRRAGMGDGTMEVEDDSDPASVYTYTVYRDGTKIKEGLTATTTFEEDGVAAG 915
DB 116 FEAGKKYTFTMRRAGMGDGTMEVEDDSDPASVYTYTVYRDGTKIKEGLTATTTFEEDGVAAG 175

QY 916 NHEVCYEVKYTAGVSPKVKDVTVEGSNEFAPVQNLTGSSVGQKVTWKWDAPNGTNPNP 975
DB 176 NHEVCYEVKYTAGVSPKVKDVTVEGSNEFAPVQNLTGSSVGQKVTWKWDAPNGTNPNP 235

QY 976 NPNPNG- ---TTLSSEFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSGCVYSESFG 1032
DB 1 ANEAKVVLAAADNVGNDTGYQFLLDADHNTFGSVIPATGTLTASSNLYSANFEYLIP 60
```

Db 236 NPENPNPNTTLESSENGIPASWKTIDADGCGHGWKGNAPGIAGYNSNGCVYSFSG 295
 QY 1033 LGGIGVLTDPNLYLTALDLPNGKLFVWCAQDANYASEHYVYASSTGNDASNFNAL 1092
 Db 296 LGGIGVLTDPNLYLTALDLPNGKLFVWCAQDANYASEHYVYASSTGNDASNFNAL 355
 QY 1093 LEETITAKGVRSPKAIIRGTOGTWRQKTVDLPAGTKVAPRHPFQSTDMFYIDLDEVEIKA 1152
 Db 356 LEETITAKGVRSPKAIIRGTOGTWRQKTVDLPAGTKVAPRHPFQSTDMFYIDLDEVEIKA 415
 QY 1153 NGKR 1156
 Db 416 NGKR 419

RESULT 6

US-10-174-695-6
 ; Sequence 6, Application US/10174695
 ; Publication No. US20030232022A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: Slakeski, Nada
 ; APPLICANT: Chen, Chao Guang
 ; APPLICANT: Barr, Ian George
 ; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
 ; FILE REFERENCE: 529282000700
 ; CURRENT APPLICATION NUMBER: US/10/174,695
 ; CURRENT FILING DATE: 2002-06-18
 ; PRIOR APPLICATION NUMBER: PCT/AU00/01588
 ; PRIOR FILING DATE: 2000-12-21
 ; PRIOR APPLICATION NUMBER: AU PQ 4859
 ; PRIOR FILING DATE: 1999-12-24
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: Fast-SEQ for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 231
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-174-695-6

Query Match 13.5%; Score 1241; DB 15; Length 231;
 Best Local Similarity 99.6%; Pred. No. 5.5e-87;
 Matches 230; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 759 FLDDADHNTGSGVLPARGPLFTGTASSNLNLSANFEYLIPANADPVVTTQNIIVTGGQEVV 818
 Db 1 FLDDADHNTGSGVLPARGPLFTGTASSNLNLSANFEYLIPANADPVVTTQNIIVTGGQEVV 60
 QY 819 IPGGVVDYCLTNPPEPAGKMWIAGDGNOPARYDDFTFEAGKKYFTTMRPAGKGDGTDME 878
 Db 61 IPGGVVDYCLTNPPEPAGKMWIAGDGNOPARYDDFTFEAGKKYFTTMRPAGKGDGTDME 120
 QY 879 VEDDSPASYTYTVYRDGDKIKEGLTATTFEEDGVAAGNHEVCYEVKVTAGVSKVCKDVT 938
 Db 121 VEDDSPASYTYTVYRDGDKIKEGLTATTFEEDGVAAGNHEVCYEVKVTAGVSKVCKDVT 180
 QY 939 VEGSNFAPVQNLTGSSVGQKVLTKWDAPNGTNPNNPNPNPNTLTSESF 989
 Db 181 VEGSNFAPVQNLTGSSVGQKVLTKWDAPNGTNPNNPNPNPNTLTSESF 231

RESULT 7

US-10-387-977-105
 ; Sequence 105, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 529282000301

; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PQ 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: Fast-SEQ for Windows Version 4.0
 ; SEQ ID NO 105
 ; LENGTH: 736
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-105

Query Match 5.9%; Score 541; DB 15; Length 736;
 Best Local Similarity 25.1%; Pred. No. 3.2e-32;
 Matches 200; Conservative 130; Mismatches 328; Indels 140; Gaps 29;
 QY 3 KLALLIA-ASLLGVGLYQAQSAKIKLDAPTTTCTNNSPKQFDASFSFNEVELTKVETKG 61
 Db 7 RIVSIWAFSSLLGMAFAQPAE-RGRNPQVRLLSAEQSMK--VQFRMDNLQFTGVGTSK 63
 QY 62 GTFASVSI PGAPFTGEVSGPEVPAVRKLIAPVPGATPVVRVKS--FTEQVYVSLNQYGEK 119
 Db 64 GVAQVPTFTGEGVNISEKGTPIILPILSRSLAVSETRAMKVEVVSKEFKKXDLV----- 116
 QY 120 LMPHQSMKSDDEPEKVPFVYNAARVAKGVFQGEQLTQVEMLGTMRGVRIAALTINPVQY 179
 Db 117 IAPSKGVISRAENPDQIPIYVYQSYNEDEKFFPGEIATLSDPF-ILDRVGRQVWVNFAPLQY 175
 QY 180 DVVANOLKVRNNIEIVSFGADEVATQRLYDASFSPPETAYKQLENRDVYTDHGLYN 239
 Db 176 NPVTILRIYTVVAVSETAEAGQNTISLVKNTFTGFEDIIYKVSFPMNTEATRY----- 230
 QY 240 TPV-----RMLVAGAKFKEALKPMLTWAKQKGFYLDVHYTDEAEVGTNTASIKAFIHK 293
 Db 231 TPVEEKENGRIWIVPKKYEEDIEDFVDMKNQGLRETVKVAEDIASPVTANAIQQPVVK 290
 QY 294 KY----NDGLASAAFPVFLALVGDTDVISGE--KGKTKVTDLYYSVAVDGFPPDEMYTF 347
 Db 291 EYKEGND-----LTYVLLVGDHDKDIPAKITPGIKSDQV---YQGIQVNDHYNEVFIG 340
 QY 348 RMSASSPEELTNIIDKVLVMEKATMPDKSYLEKVLIIAGADYSWNSQVGOPTIKY-GMQY 406
 Db 341 RFSCSKEDLKTQIDTIHYERNITTEDKWLGGALCIASAEGPSADNGESDIQHENIIA 400
 QY 407 YNQEHGYTDVYNYLKAPYTCY-----SHLNTGVSFANYTAHGETAWADPLLT 456
 Db 401 NLLTQXGYTKIIR-----CYDPGVTPKNIIDAFNGGISLANYTGHGSETAWGTSHFG 452
 QY 457 TSOLKALTNNKQKFLAIGNCCITAOQPDYVQPCGEVITRV----KEKGAYVIGSSPNSY 512
 Db 453 TTHVKQLTNSNQLPFTFDVACVNGDFLYNVPCCAEALMRAQKDGKPTGTVAIIASTINQS 512
 QY 513 WGEDYVWSGANAVFGVQPTFEQTSMSGYDATPLEDSYNTVNSIMWAGNLAATHAGNI-- 570
 Db 513 W-----ASPMRG-----QDEMEI-----LCEKHPNIIKR 537
 QY 571 --GNITHIGAHYWEAYH-----VLGDGSMYPYRAMPKNTNTYTLPLASLPQNAS 617
 Db 538 TFGGVTMNGMFAMVEKYKQGEKMLDWTWVFGDPSLLVRLTPTKMQVTPANISASAO 597
 QY 618 YSIQAS-AGSYVAISKDGVLYGTGVANAGVATVSMTKOITENGNDVVDVITRSNYLPVIK 676
 Db 598 FEVACDNGAIATLSDDGDMVGTAIVK-DGKAIILKLNESIADETNLTLTVVGVNKKVTVIK 656
 QY 677 QIQVGEPSPYQVPSN---LTATTQGGQKVLTKWEAPSAKKAEGSREVKRGIDGLFVTIEPA 733
 Db 657 DVKV-EGTSIADVANDKPTVAVSGKTIIV---ESPAA-----GL---TIFDM 697
 QY 734 NDVRANEAK--VVLADN 749

Db 710 PNTPVATPIVETVGTSTTKGIEV---KGTAEVGTITIEVRDAAG-----TVLGTATTGTD 760
QY 901 GLTATTFFEDGVAAGNHEVCVEVKYTAGVSPKVC-----KQVTVGSNEFAP-VQNLGTS 954
Db 761 GXYTVTL-DSGTATANQILSVAKNASGTESQATATTPADV-----APTVDNITGN 812
QY 955 S-VQKVLTKWDAPNGTPNP-----PNNPMPGPTTL 985
Db 813 SGSGYEIT-----GTADENTTIEVRDPGAVIGTSDANGDFTVTLPGTINPGDGL 865
QY 986 SESFENGIPASWKTTIDADGCHWPG-----NAPGTAGVNSGCVSEFGLGG 1035
Db 866 T-----VTKGNAGNESQFTEVLVPADATVAPTQVGTGNSVA----- 904
QY 1036 IGVLPDNYLIPALDLPNGGKLTFWCAQDANYASEHVAVASSTGNDASFT----- 1089
Db 905 -----GQVGTGAD-PNA-----TIEIRDAGN-----VIATGATGTSFAVNLPA 946
QY 1090 NALLEETITAKVRSFKAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEV- 1148
Db 947 TANANETLTA-----LAKDPAG-----NTSTPTTFQTPADEV 979
QY 1149 -----ELKANGKRA-DTEFESSTHGEAPAEWTI-----DADGQGWLCSSGQDWM 1197
Db 980 APPSVDKVTGNTQGYQVGTGAEGLTTEIEVRATDGTGVLGTATTGPTGYTVTLASGK--- 1036
QY 1198 LTAHGSNVSVSFGSMALPNPNLYLISKDVGTGATKVKYVAVNDGFPDGHVAVMISKTG 1257
Db 1037 ATAKQTVAVAK-NDTGLESQPTTAMTADVTPT-----LAKDPAG-----NTSTPTTFQTPADEV 1072
QY 1258 TNAGDFTVFEET-----PNGI-----NKGARPLSTEANGAKPQSVWIERVDPAGPKY 1309
Db 1073 DITGDSITGYEITGADENTTIEVRNPDPGTIIGTTTDDQG-----NFTVDPAGAA- 1124
QY 1310 VAFRHVNSDLNYILLDDIQTMTGSSPTDYV-----TVYRDGKTKIEGTEITFEEDG 1365
Db 1125 -----NFGDLTVVKGDD-----GNESQTEVTVPEDAVAPTQVTVGTATGYQVGT 1175
QY 1366 VATGNHEVCVEVKYTAG-VSPKCKVDVTNQTQFNPQNLTAEOAPNSMDAILK----- 1418
Db 1176 TAEPN--VTIEHNEAGLVATGTTDAGAFITLPTGTATANE-----LTAIAKDAAGKE 1230
QY 1419 -----WNAKRAEVLNEDFE-----NG-----IPASWKTTIDAAG----- 1449
Db 1231 SNPTAFKTPADPAVAPTPTVDKITGTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTG 1290
QY 1450 -DGNWTTTPPPGGSSPAGHNSAICVSSASHNFEQPQNPDLVTP-ELSLPGGTLTF 1507
Db 1291 TDG-KYTVTLFPGKAS---ANETIVVAK---NATGKESQATATTPVDLATPTIDSIT- 1342
QY 1508 WVAQDANYASEHY-----AVYASSTGNDASNFANALLEEVL- 1547
Db 1343 -----GNSKGYEITGTAEPKTTIDVRDADGTIIAATTANETQYTVTLPAQVTFGE 1395
QY 1548 TVVTAPEAIRCTAGGTWYQKTVOLPACTKTVAFRHFCTDFFWINLDDVWITSGNAPSY 1607
Db 1396 TITISKDGAGNESQ---PATAVIPA-----DVVLA--AP-- 1425
QY 1608 TYTIYRNNTQIASGVTTETTRDPOLATGYTYGVKVYVNGESALETATLNTSLADVTA 1667
Db 1426 --TITKVEGNKANGYTVGTADPNVTQVFNSSQLLASGNTTGGTFVHIA--AGLAT 1481
QY 1668 QKPYTLTVGKTIITVTCQ 1686
Db 1482 EKE-TLT-----ALATDTQ 1495

RESULT 12

US-10-282-122A-47453
; Sequence 47453, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47453
; LENGTH: 2435
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47453

Query Match 3.38; Score 303.5; DB 12; Length 2435;
Best Local Similarity 21.58; Pred. No. 4.8e-13;
Matches 309; Conservative 121; Mismatches 538; Indels 467; Gaps 68;

QY 531 PTFEGTS-----MGSYDATELEDSYNTVNSIMW-----AGNLAATHAGNIGNIT 574
Db 244 PTLGGTAAGSRVSVYDGTLLGTTTADSSGKWTFTTTLGEGAHSTVTTATDAAGNV 303
QY 575 HIGHYWEAHVHLGDGSMVPYRAMPKNTYITPASPQNOA-----SYSIOASAGSYVAI 630
Db 304 -----TPSAAFELTIDTTAPA-LPTVNAITDGTSLSGTAAGATVNI 343
QY 631 --SKDGVLYGTGVANASGVATVSMTKQITENGVDVYITRSNYLPVKIQVGEPSYQP 688
Db 344 DTNGDGTDAITVADPSGWT-----YTP-----STPLDGT 375
QY 689 VSNLTAT-----TQGQKVTLKWEAPSAK-----KAEGSREVKRIGDG-----L 726
Db 376 VIGVTATDAAGNTGPSASVITVGTDTTAPGAPVIGTVDAGSVVGAIASGSDTDDATP 435
QY 727 FVIEPANDYRANEKVUL-----AADNVWGDN 754
Db 436 SGTAAGSTVSVYDGTLLGTTTADPSGNWTTFTTTLGEGAHSLATVTTATDAAGNV 495
QY 755 TGQOFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEVLVPANADPVVTTQNIIVT 814
Db 496 TAPDLTIDITAPSIPTVNTADGTSLSGTABAG-----ATINIDTNGD 537
QY 815 GE-----VVIPGGVYDYCIINPEPA-----SGKMWIAGDGGNQPARYD 852

```
Db 538 GTPDAITVADPGAWYITSTPLPAGTVIGVATDAAGNTGPSASVTVTGDTTAPGAPVI 597
QY 853 DTFEAGKYYTMRRAAGMGDTMVEDSDPSASYTYTVVROGTKIKEGLTATTFEEDG- 911
Db 598 GTVDDAGSVVGAISGGSTDDATPTLSGTAEAGSTSVY-DGTTLL---LGTTADPSGN 653
QY 912 -----VAAGNHECYEVKVTAG-VS-PVKCKDVTVEGSNEFAPVQNL-T-GSSV--- 956
Db 654 WFTPTTALGEGASHLTVATDTAGNVSPSTAFDITITDTAPAITVNAVATGTSUSGTA 713
QY 957 --GQKVLTKWDAPNGTPNPNPNPN-----PGTTLSEFENGIPASWKTIDADGD--- 1005
Db 714 EAGATVNIWDNG-DGTPDATVADPGAWYITSTPLPAGTVIGVAT---DAAGNTGP 768
QY 1006 -----GHGWKPG-----NAPGIAGYNSNGCVYSSEFSG 1032
Db 769 SASVTVTGTTAPGAPVIGTVDDVGVVGAIISSGSTDATPTLSGTAEAGSTSVYDG 828
QY 1033 LGGIGVLTDP---NYLITPALDLENGG---KLTFWCAODANVASEHYAVASST----- 1081
Db 829 TLLGTTTADPSGNWTFPTTGLGEGASHLTVATDTAGNVSPSTAFELTIDTTAPALP 888
QY 1082 -----GNDASFTNALLEETITAKGVRSFKA-IRGRIQGTWR-QKTVDLPAGTKYVAFR 1133
Db 889 TVNATDGTSLSGTAEGATVNIWDNGDGTDPATVADPGAWYITSTPLPAGT----- 942
QY 1134 HFQSDTMFVIDLDEVEIKAN-KRADFTETFSSTHGEAPAEWTTIDADGCGGWCLCS 1192
Db 943 -----VIGVATDAAGNTGPSASVTVTGDTTAFG-APVIGTVDDVGVVGAII-SG 992
QY 1193 GOLDMLE-----TAHGGSNVVSFSKNGMAL-----NPDNYLISKDVTGATKVKYXAV 1240
Db 993 GSTDDATPTLSGTAEAGSTV---SVVDGTYLLGTTTADPSGNWTFPTTGLGEGASHLV 1049
QY 1241 -----NDGFFGDHYAVIMSKTG-----TNAGDFTVVFETPNKGGARFGLSTEAN 1288
Db 1050 TATDAAGNVSTSTAFDITIDTTAPALPTVATDGTSL-----SGTAEGATVNIWDNG 1104
QY 1289 G-----AKPOSVW-IERTVDLPAGT-----DYTVTVVROGTKIKEGLTETTFEEDG- 1365
Db 1105 GTPDAITVADPGAWYITSTPLPAGTVIGVATDAAGNTGPSASVTVTGDTTAPAAPIV 1164
QY 1322 YTLDDIQITMG-----GSPPT-----DYTVTVVROGTKIKEGLTETTFEEDG- 1365
Db 1165 VSVTDDVGSIVGLTGTATDDTTPLAGTAEGSTVNY-DSTLL---LGTTADPSGN 1220
QY 1366 -----VATGNHECYEVKVTAGVSPKCKVDVTVNSTQFNPVQNLTAEQAP--NSMDA 1415
Db 1221 WFTPTTALGEGASHLTVATDTAG-----NVSPSTAFNLITDTTAPALPTVNAVTD- 1272
QY 1416 ILKNWAPASKRAEVLNEDFENGIPASWKTIDADGDN-NWTTTPPGSGSFAGHNSAICV 1474
Db 1273 -----GTSLSGTAE-----GATVNIWDNGDGTDPATVADPSGA----- 1307
QY 1475 SSASHINFRGPQNDNYLVTPELSPGGGTLTFWVCAQDANVASEHYAVASSTGNDASN 1534
Db 1308 -----WYTPSTPLPAGTVI--GVATATA-----AGNTGPSAS- 1338
QY 1535 FANALLEVLTAKTVVTAPEA-IRGTRAQGTWYQKTVQLPAGKYVAFRHFQCTDFFWIN 1593
Db 1339 -----VTVGTDTTAPAAPIIIG-----TVTDDVGSWGAIISGGSTD----- 1374
QY 1594 LDDVWITSNAPS-YVYTYIRNNQIAGSVTETTYRDPD-----LATGFYTYGVK 1642
Db 1375 -DATPTLSGTAEAGSTSVVDGTTLLG-----TTTADPSGNWTFPTTALGEGASHLTV 1428
QY 1643 VVYPNGESAETATLNTISLADVTACKPVTLTVV-GKTIIVTCQGEAMI-YDMNG 1695
Db 1429 ATDTAGNVSPSTAFNVT--IDTTAPAITVNAVATDGTSLSGTAEGATVNIWDNG 1481
```

```
US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66335
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-10-282-122A-66335

Query Match 3.3%; Score 300.5; DB 12; Length 2468;
Best Local Similarity 20.5%; Pred. No. 8.4e-13;
Matches 305; Conservative 140; Mismatches 505; Indels 535; Gaps 74;

QY 534 EGTSMGSDYATFLDPSYNTVNSIMMAGNLAATHAGN-----IGNITHIGAH 579
Db 99 DGPLMAGY---IPQESLAGFESLTGAGVLGMSAGTALLVGAAGIAGVAISNSSGGGG 155
QY 580 YWEAYHVLGDSVMPYRAMPKTYTTLPASLPQNASYSIQASAGSYVAI---SKDGVLY 637
Db 156 -----GGSSVFPDDTTTPPKAAS---GLKLPADGGSSISGQAAGASVGDITNGDKPD 203
QY 638 GTGVANASGAVTVSMTKOITENGNDVWIT--RSNVLPIVKIQVCEPSPYQPVSNLTAT 695
Db 204 LTVIADANGNFTAPINPPLNTNGTQTVTVVTDPAGNASPP-AQTAPDTPAPADTVQVA 262
QY 696 TQGQKVTLLKWEAPSAKKAEGSREVKRIGDG---LFVTIEPANDVRA-----NEAKVY 744
Db 263 PDGSSVTGKAEPGTVGVVDTD-----GDGQPDTTVVVGPGGSFEVPLNPLTNGETVTV 316
QY 745 LAADNVWGNTGYQFLLDADHNTFGSVIPATGPLETGTGTASSNLYSANFEYLVPANADPVV 804
Db 317 IVTDPAGNNST--PVTVEAPDPTT--APAPATDVQVAPDGSS-----VTGNAEPGA 362
```

QY 805 TTONIIVTGOGE-----VVIPGGVYDICTNPEPASGK---MWIAGDGGNQ-----PARYD 852
Db 363 TV-GVDTDGDGPDTTVVVGPGSGFEVPL-NPPLTNGETVTVITDPAAGNSFTPTAEAP 420
QY 853 DF-----TFEAGKYYFTMRRAGMGDTM-EVEDDSPASYTTVTVYRDG 895
Db 421 DFPDAPQVNASGVLGTAAGVTVIT-----DGNPNIGQTSADANGWSFT---PG 472
QY 896 TKIKEGLTATTFEEDGVAAGNHCYEVKYYTAGVSPKVKDVTVEGSENEFAPV-----Q 949
Db 473 SOLPDGTVVNVVARD--AAGNS-----SP--ATSIIVDGVAPNAPVVEPNSGS 516
QY 950 NLTGSS-VGQKVTL-----KWDAPNGTNPN----- 974
Db 517 ELSGTAEPGSSVTLTDGNGNPIGQTTADANGNSFTPTPLDGTVVNVVARDAGNSSP 576
QY 975 -----PNPNPNTLTLSSEFE-----NGIPASWKTIDADGCHGW-- 1009
Db 577 PASVTVDAVAPATPTVDFSPNGTTLGTAEPGSSVTLTDGNGNPIG--QVTDAGSGN-WTF 633
QY 1010 -----KPNAPGAGYNSGCVYS 1028
Db 634 TPSTPLENGTVVNTATDPSGNASSPASVTVDAVAPATPVVNPNGTTLGTAEPGATVT 693
QY 1029 ESFGLGG-IGVLPD-----NYLITPALDLPNGGKLTFFVCAQADANYASEHYAVYASSTGND 1084
Db 694 LTDGNGNPIGQVTDAGSGNSFTTPTLPG-----TVVNTATD 733
QY 1085 ASNFTNALLETI-----TAKGVRSKPAIRGRIQGTWR---OKTVLDLPA 1125
Db 734 ASGNTSAGSSVTVDSVAPATPVVNPNGTTLGTAEPGSSVTLTDGNGNPIGQVTD--- 790
QY 1126 GTKYVAFRHFQSTDMFYIDLDEBIKANGKRAEDTETEFES-STHGEAPAEWTTIDADG 1184
Db 791 GSGNWSFT--PSTEL-----ADGTVVNTATDPAAGTSGQGS---TIVDGVAPT 834
QY 1185 QGMCLSSGQLDWLTAGGNNV-----SFSW-----NGMALNPD 1220
Db 835 TPTVNLSSGSLGTAEPGSSVTLTDGNGNPVIAEVTADGSGNWTPTSTPIANGTVVNV-- 892
QY 1221 NYLSKQVTKATKYYAVNDGPPGDHYAVMISKTGTNAGDFTVVFEEETENGKKGAR 1280
Db 893 ---VVAQDAAGNSPGASVTVDSQAPA--APVNPNSGT-----TLGTAEPGAT 937
QY 1281 FGLSTEANG-----AKQSVWIERTVDLDPAGTKYVAFRHYNCSDLYLLDDIOFTMG 1333
Db 938 VTL-TDNGNGNPIGQVTDAGSGNV-----SFTPGTPLA-----NGTVVNTATSDPTGNTSA 986
QY 1334 GSPTPTDYTVY-----RQGTKI-----KEGLTETTFEEDG-----VATGNHCYCE 1376
Db 987 PASTTVDSVAPAAVVPNPSNGAELSGTAEPGATVTLTDGNGNPVIGQVTDAGSGNSFTPT 1046
QY 1377 VKYTAGVSPKKCVDTVNSTOFNPQNLITABQAPNMSDAILKMNAPAKRAEVLNEDPEN 1436
Db 1047 TPLADG-----TVVNTATDPAGN-TGQGGSTTVDAI---APATPTVNLSSGSL 1093
QY 1437 GIPASWKT-IDADGNG-----NWTTP----- 1458
Db 1094 GTAEPGSSVTLTDGNGNPVIAEVTADGSGNWTPTSTPIANGTVVNVVARDAGNSSPPT 1153
QY 1459 -----PG-----GSSSEFAGNSAICVSSASHNFPSPQNPQNYLITBELSL 1499
Db 1154 TVDSSAPPAPVNPNSGVLISGTAEGATVTLTDAGNPVIGQVTDAGSGNSFTPTPTPL 1213
QY 1500 PGGGTLTFVWCAQADANYASEHYAVYASSTGNDASNFANALLEEVLITAKTVVTAPEAI--- 1556
Db 1214 ANGTIVIV-----ATADPTGNTGQAA-----TTVDAVAPAPVIDP 1250
QY 1557 -RGTRAQGTWQVQLPAGTKYVAFRHFCTDFWNLDDVITSGNAPSFTYTI----- 1611
Db 1251 SNGTTSIGT-----AEAGAK-----VILTLDGNGNPIGETTAGDS 1284
QY 1612 ----YRNNITQIASG-VTETTYRDPDLATGFTYGVKVVYPNGESALETATLNIIT----- 1660

Db 1285 GNSFTPGTFLANGTVVNAVAQDPAGNTG-----PQGSTTVDAVAPNTFVVPNS 1334
QY 1661 -SLADVTAQKPYTLTVVGKTIITVTCQGEAMLYDMNGRRLAAGNT 1704
Db 1335 GNLLNGTAEP-----GSTVILT-----DGNGNPFI--GQTT 1362
RESULT 14
US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4
Query Match 3.3%; Score 300.5; DB 14; Length 2468;
Best Local Similarity 20.5%; Pred. No. 8.4e-13;
Matches 305; Conservative 140; Mismatches 505; Indels 535; Gaps 74;
QY 534 ECTSMGSDATPFLSDSVNTVNSIMWAGNLAATHAGN-----IGNITHIGAH 579
Db 99 DGPLMAGY-----IPQESLAGFESLTGAGVLGMSAGTALLVGAAGAGVAINSSGGGG 155
QY 580 YYEAVHVLGDGSMVPYRAMPKNTVTLTPASLPQNOASYSIQASAGSYVAI--SKQGVLY 637
Db 156 -----GGSSVPPDTPPKAAS---GLKIAPDGSSISGQAEGASVUIDINGDKPD 203
QY 638 GTGVANAGVATVSMTKQITENGNYDVVIT--RSNYLPVTKIQVGEPSYPQVSNLTAT 695
Db 204 LTVIADANGNFTAPLNPPLTNGQTVTVVWTDPAAGNASPP-AQVTAPDTPATADYQVA 262
QY 696 TQOKVTLKWEAPSAKKAEGSRVKRIGDG--LFTVIEPANDVRA-----NEAKVY 744
Db 263 PDGSSVTGKAEPSGTVGVDTD-----GDCQPDPTVVVGGGSEFEVPLNPLTNGETTV 316
QY 745 LAADNVMGDNTGYQLLADADHNTFGSVIPATGFLFTGTASSNLYSANFEYLVPANADPVV 804
Db 317 IVTDPAGNNST--PVTVEAPDPTT--APAPATDVQVAPDGSS-----VTGNAEPGA 362
QY 805 TTONIIVTGOGE-----VVIPGGVYDICTNPEPASGK---MWIAGDGGNQ-----PARYD 852
Db 363 TV-GVDTDGDGPDTTVVVGPGSGFEVPL-NPPLTNGETVTVITDPAAGNSFTPTAEAP 420
QY 853 DF-----TPEAGKYYFTMRRAGMGDTM-EVEDDSPASYTTVTVYRDG 895
Db 421 DFPDAPQVNASGVLGTAAGVTVIT-----DGNPNIGQTSADANGWSFT---PG 472
QY 896 TKIKEGLTATTFEEDGVAAGNHCYEVKYYTAGVSPKVKDVTVEGSENEFAPV-----Q 949
Db 473 SOLPDGTVVNVVARD--AAGNS-----SP--ATSIIVDGVAPNAPVVEPNSGS 516
QY 950 NLTGSS-VGQKVTL-----KWDAPNGTNPN----- 974
Db 517 ELSGTAEPGSSVTLTDGNGNPIGQTTADANGNSFTPTPLDGTVVNVVARDAGNSSP 576
QY 975 -----PNPNPNTLTLSSEFE-----NGIPASWKTIDADGCHGW-- 1009
Db 577 PASVTVDAVAPATPTVDFSPNGTTLGTAEPGSSVTLTDGNGNPIG--QVTDAGSGN-WTF 633

QY 1010 -----KPNAPGIAGYNSGCVYS 1028
Db 634 TSTPLPNGTVVNNATDPGNSASSPASVTVDAVAPATPVVNSGTTISGTAEPGATVT 693
QY 1029 ESGLGG-IGVLFPD---NYLIITPDLPLNGGKLTFWVCAQDANYASEHYAVASGND 1084
Db 694 LTDGNGNPIQVTDGSGNWSFTPTPLPNG-----TVVNTATD 733
QY 1085 ASNFTNALLEETI-----TAKGVRSPKAIIRGLQGTWR---QKTVDLPA 1125
Db 734 ASGNTSAGSVTVDSVAPATPVINPNNGTTLSTGABGSSVTLTDGNGNPIQVTD--- 790
QY 1126 GTKYVAFRHFQSDMFYIDLDEVEIKANGKRAOFTETFEES-STHGAPAEWTTIDADGG 1184
Db 791 GSGNWSFT--PSTPL-----ADGTVVNTATDPAGTSGGGS-----TTVDGVAPT 834
QY 1185 QGWLCSGGLDMLTAHGGSNV-----SSFSW-----NGMALNPD 1220
Db 835 TPTVNLNSGSLSGTAEPGSGTVILTDGNGNPIAEVTDAGSGNWTYPTSTPIANGTVVN-- 892
QY 1221 NYLISXDVTCATKVKYXYAVNDGFGPDHYAVMISKTGTNAGDFTVVFEETPENGINKGAR 1280
Db 893 --VVAQDAAGNSPGASVTVDSQAPA--APVVPNSGT-----TSLGTABPGAT 937
QY 1281 FGLSTEANG-----AKQSVMIERTVDLPAGTKYVAFRHYNCSDILYLLDDIOFTMG 1333
Db 938 VTL-TDNGNPIQVTDAGSGNW-----SFTPGTPLA-----NGTVVNTASDPTGNTSA 986
QY 1334 GSPTPTDYTVVY-----RDGKI-----KEGLTETTFEEDG-----VATGNHEYCV 1376
Db 987 PASTTVDSVAPAPVVPNSGABISGTAEPGATVTLTDGNGNPIQVTDAGSGNWSFTPS 1046
QY 1377 VKYTAGVSPKKVDVTVNSQFQVONLTAEQPNMSDAILKXNAPASKRAEVLNEDFEN 1436
Db 1047 TPLADG-----TVVNTATDPAGN-TGGQGSTTVDAI-----APATPTVNLNSGSL 1093
QY 1437 GIPASWKT-IDADGDN-----NWTTP-----1458
Db 1094 GTAEPGSTVILTDGNGNPIAEVTDAGSGNWTYPTSTPIANGTVVNVVAQDAGNSPPAT 1153
QY 1459 -----PP-----GSSSPAGHNSAICVSASHINPEGQPNQPNLYVTPELS 1499
Db 1154 VTVDSSAPPAPVVPNSGNGVVISGTAEPGATVTLTDGNGNPIQVTDAGSGNWSFTPTPL 1213
QY 1500 PGGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANALLEVLTAKTVTVAPEAL--- 1556
Db 1214 ANGTIVV-----ATATDPTGNTGQAA-----TTVDVAVAPPAPVIDP 1250
QY 1557 -RGTRAQGTWYQKTVOLPAGTKYVAFRHFCTDFFMINLDDVVITSGNAPSYYTYI--- 1611
Db 1251 SNGTTISGT-----AEAGAK-----VILTDGNGNPIGETTADGS 1284
QY 1612 -----YRNNTQIASG-VTETTYRDPDLATGYTYGVKVYVYNGESAETATLNT--- 1660
Db 1285 GNWSFTPGTFLANGTVVNAQDPAGNTG-----PQGSTTVDAVAPTPVVPNSN 1334
QY 1661 -SLADVTAQKPYLTVVKGKTIYVTCQGEAMIDMNGRRLAAGRNT 1704
Db 1335 GNLLNGTAEP-----GSTVILT-----DONGNPI---GQTT 1362

RESULT 15

US-10-206-576-258
; Sequence 258, Application US/10206576
; Publication No. US20030017495A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/206,576
FILING DATE: 29-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/071,035
FILING DATE: 1998-05-04
APPLICATION NUMBER: US 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: US 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: US 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB369PD1
INFORMATION FOR SEQ ID NO: 258:
SEQUENCE CHARACTERISTICS:
LENGTH: 1638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 258:
US-10-206-576-258

Query Match 3.2%; Score 293; DB 12; Length 1638;
Best Local Similarity 21.5%; Pred. No. 1.6e-12;
Matches 293; Conservative 140; Mismatches 499; Indels 430; Gaps 72;
QY 520 SVGANAVGQPTFGTSGSYDATEFDSYNTVNSIMMAGNLAATHAGNIGNITHIGH 579
Db 279 AVGGS---GVNELVDASLLGTTT-TPTVSPQNL---SNNLDARFVGTVVQTDLLDN 332
QY 580 YWEAHHVLGDGSMVYRAMPKNTVTLTLP---ASLPQNOASYSIQASAGSYVAI---SKD 633
Db 333 LIATA-----DGVSNIFYAAGTTSEVTAPTITGVNTAGYEVKGTADANATEIRNAG 387
QY 634 GVLVGTGVANASGATVSMTKQITENGYDVIITRSNYLFPVIKQIQVGE--PSPYQ-PVS 690
Db 388 GTVIGTGTADGTGAFV--TVPAGEAGANETLTA-----VAKNASGTEXTPTTFTQTPAD 439
QY 691 NLTA-----TQGVKVTLKWEAPSAKKAEGSREVRI-----GDGLFVIEPA 733
Db 440 EATVTAPTITGVTVGNS-TAGYEVKGTADANATEIRNAGGTVIGTGTADGTGFTVTPA 498
QY 734 NDVRANEAKVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPTGTASSNLYSANPE 793
Db 499 GEAGANETLTAVAKNASGTSTPTTFTQTPADEAT-----VTAPTITGVTVGNS--TAGYE 550
QY 794 YLVPANADPVVTTQN-----IIVTQGEVVIQGVVDYDITNPEPASGKMWIADGG--- 845
Db 551 VKGTADANATEIRNAGGAVIGTGTAD-----GTGFTVTVIP-----AGEAGANET 596
QY 846 -----NOPAR---YDDFTFEAGKKYFTTMRAGMCD-GTDM 877
Db 597 LTAVAKNASGTSTPTTFTQTPADNTPVATPVTETVGTSTTKGYEV-----KGTAEVGTTI 652
QY 878 EVEDDDSPASYTVTVYRDGKIKKGLTATTFEEDGVAAAGNHEVCVEVKYTAGVSPKVC--- 934
Db 653 EVRDAAG-----TVLGTATGTGDKYTVTL-DSGTATANTQTLSSVAKNASGTESQATAT 706
QY 935 --KDVTVGSGNEFAP-VQNLGTSS-VGQKVTWKWDAPNGTNPFN-----974

```
Db 707 TPADVT-----APTVDNITGSGGYEIT-----GTADENTTIEVRDPSGAVIGT 752
Qy 975 -----PNPNPNPGTTLSEFENGIPASWKIIDADGHHGKPG-----1012
Db 753 TSDANGDFTVTLPTGTTNPGDILT-----VIGKNAGNESOPTTEVLVPADATV 800
Qy 1013 NAPGIAGYNGGCVYSSFGLGIGVLTPDNYLITPALDLPNGGKLFVWCAODANYASE 1072
Db 801 TAPTVTGVTGNSVA-----GYQVTGTAD-PNA---TIERDADGN---836
Qy 1073 HYAVYASSTGNDASNFT-----NALLEETITAKGVSRPKAIRGRIQGTWRQKTVDLDPAG 1126
Db 837 ---VIATGTADGTGSPAVNLPAGTANANETLTA-----LAKDPAG 873
Qy 1127 TKYVAFRHFQSTDMFYIDLDEV-----ELKANGKRA-DETFETFESETHGEAPAEWTI 1178
Db 874 -----NTSTPTTFQTPADEVAVAPRVDKVTGNTQGYQVTGFAELGTTIEVRATDGTV 926
Qy 1179 ----DADGDGGWMLCLSSGOLDMLTAHGGSNVSSFGNGMALNPDNYLISKDVTGATKV 1234
Db 927 LGIATATTGPTGOYTVTLASGK---ATAKQTVNVVAK-NDTGLESOPTTAMTADVTITPT--980
Qy 1235 KYIYAVNDGPPGDHYAVMISKTGNAGDFTVVFEET-----PNGI-----NKGARFGLSTE 1286
Db 981 -----IGDITGDSITGYEITGTADPNTTIEVRNPDGTIIGTTTT 1019
Qy 1287 ANGAKPOSVMIERVLDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGSSPTPTDYTY---1343
Db 1020 DDQG-----NFTVDLPAGAA-----NPGDILTUVGKGDG---GNESOPTTEVTVPED 1062
Qy 1344 -TVVRDGTIKIKGLTETTFEEDGVATGNHEYCEVVKYTAG-VSPKCKVDVTNSTQFNVPV 1401
Db 1063 ATVAAPTVTVTGTTATGYQVTGTAEPN--VTIEIHNEAGLVIATGTTDGAGAFITILPT 1120
Qy 1402 QNLTAEOAPNSMDAILK-----WNAPASKRAEVLNEDFE-----NG-----1437
Db 1121 GTATANEA---LTAIAKDAAGKESNPTAFKTPADPDAPFATPTVDKITGTTNGYQVVG 1177
Qy 1438 --IPASWKTIDADG-----DGNWTTTTPPGSSPAGHNSAICVSSASHINFEGR 1485
Db 1178 AEVGTIEVRDADGTVLGMATGTGTDG-KYTVTLEFGKAS---ANETITVVAK---NATGK 1230
Qy 1486 QNPDNYLWTP-ELSLPGGGTILTFWVCAODANYASEHY-----AVYA 1525
Db 1231 ESQPATATTPVDLATPTIDSIT-----GNSSKGYEITGTABPKTTIDVRDADGTIIA 1282
Qy 1526 SSTGNDASNANALLEEVLIT-AKTVVTAPEAIRGTRAQGTWYQKTVQLPAGTKYVAFRHF 1584
Db 1283 ATTANETGQYTVTLFAGVTPGETITIIISKDAGNESQ---PATAVIPA-----1328
Qy 1585 GCTDFFWINLDDVWITSGNAPSYYTYIYRNNTQIASGVTEYTYRDPDLATGFYTYGVKVV 1644
Db 1329 -----DVLIA---AP---TITKVEGNKANGYTVGTADPNVTVOFYNSSEQLL 1370
Qy 1645 YPNGESAIETATNLITSLADYTAOKPYTLTVVGKTIITVTCQ 1686
Db 1371 ASGNNTTGGTFPSVHIA--AGLATEKE-TLT-----ALTTDTQ 1405
```

Search completed: May 18, 2004, 11:38:32
Job time : 74.1557 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:33:39 ; Search time 51.2572 Seconds
(without alignments)
2739.638 Million cell updates/sec

Title: US-08-570-311-2

Perfect score: 2641

Sequence: 1 MRKLNLSFSAVLVLSLLCWG.....QNLTGSVAGKVTLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: , 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 2641 | 100.0 | 497 | 2 AAR96025 | Aar96025 P. gingiv |
| 2 | 2641 | 100.0 | 497 | 2 AAW69483 | Aaw69483 Haemagglu |
| 3 | 2641 | 100.0 | 2628 | 2 AAR96030 | Aar96030 P. gingiv |
| 4 | 2641 | 100.0 | 2628 | 2 AAW69488 | Aaw69488 Haemagglu |
| 5 | 951.5 | 36.0 | 1706 | 2 AAW24786 | Aaw24786 PrtR anti |
| 6 | 944.5 | 35.8 | 1704 | 2 AAR70188 | Aar70188 Arg-gingi |
| 7 | 944.5 | 35.8 | 1704 | 2 AAW34843 | Aaw34843 Arg-gingi |
| 8 | 944.5 | 35.8 | 1704 | 3 AAY67396 | Aay67396 Arg-gingi |
| 9 | 943.5 | 35.7 | 1687 | 2 AAR08938 | Aar08938 P. gingiv |
| 10 | 943.5 | 35.7 | 1687 | 2 AAR96033 | Aar96033 P. gingiv |
| 11 | 943.5 | 35.7 | 1687 | 2 AAW69495 | Aaw69495 Haemagglu |
| 12 | 816 | 30.9 | 1358 | 2 AAR96032 | Aar96032 P. gingiv |
| 13 | 816 | 30.9 | 1358 | 2 AAW69494 | Aaw69494 Haemagglu |
| 14 | 812.5 | 30.8 | 1732 | 2 AAR96029 | Aar96029 P. gingiv |
| 15 | 812.5 | 30.8 | 1732 | 2 AAW24787 | Aaw24787 PrtK anti |
| 16 | 812.5 | 30.8 | 1732 | 2 AAW69487 | Aaw69487 Haemagglu |
| 17 | 806.5 | 30.5 | 1087 | 2 AAR96028 | Aar96028 P. gingiv |
| 18 | 806.5 | 30.5 | 1087 | 2 AAW69486 | Aaw69486 Haemagglu |
| 19 | 791.5 | 30.0 | 419 | 4 AAR03572 | Aar03572 P. gingiv |
| 20 | 662.5 | 25.1 | 419 | 4 AAU03574 | AAu03574 P. gingiv |
| 21 | 546.5 | 20.7 | 450 | 2 AAR96021 | Aar96021 P. gingiv |
| 22 | 546.5 | 20.7 | 450 | 2 AAW69489 | Aaw69489 Haemagglu |
| 23 | 546.5 | 20.7 | 456 | 2 AAR96023 | Aar96023 P. gingiv |
| 24 | 546.5 | 20.7 | 456 | 2 AAR96022 | Aar96022 P. gingiv |
| 25 | 546.5 | 20.7 | 456 | 2 AAW69491 | Aaw69491 Haemagglu |

ALIGNMENTS

RESULT 1

AAR96025
ID AAR96025 standard; protein; 497 AA.

XX AC AAR96025;

XX XX
DT 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)

XX P. gingivalis haemagglutinin hgaA.
XX XX

XX Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.
XX XX

XX OS Porphyromonas gingivalis; strain 381.
XX XX

XX PN W09617936-A2.
XX XX

XX PD 13-JUN-1996.
XX XX

XX PF 11-DEC-1995; 95WO-US016108.
XX XX

XX PR 09-DEC-1994; 94US-00353485.
XX XX

XX PA (UYFL) UNIV FLORIDA.
XX PA (UABR-) UAB RES FOUND.

XX PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX XX

XX DR WPI; 1996-287181/29.
XX DR N-PSDB; AAT30649.

XX XX
PT Porphyromonas gingivalis genes and proteins - used in the detection and
XX XX vaccination against periodontal disease.
XX PS Claim 5; Page 47-51; 153pp; English.

XX CC A portion (AAR96025) of P. gingivalis 381 haemagglutinin hgaA (see also
XX CC AAR96030) was identified as the product of clone ST2 (AAT30649) isolated
XX CC from a P. gingivalis 318 genomic library. The haemagglutinin can be obtd.
XX CC animals against periodontal disease. Expression in Salmonella cells
XX CC from transformed host cells and used as a vaccine to protect humans or
XX CC allows prodn. of a live vaccine. The haemagglutinin can also be used to
XX CC detect the presence of anti-P. gingivalis antibodies and to raise
XX CC monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-2003 to
XX CC standardise OS field)

XX SQ Sequence 497 AA;

Query Match 100.0%; Score 2641; DB 2; Length 497;
 Best Local Similarity 100.0%; Pred. No. 2.5e-194;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLSLFLAVLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60
 DB 1 MRKLSLFLAVLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60

QY 61 ARIILEAHVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEYKVPVNA 120
 DB 61 ARIILEAHVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEYKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

QY 181 VORQSGDAAASVVVTGEGNEFAPVQNLQWSVSGTQVTLTWQAPASDKRTYVLYNESFTQ 240
 DB 181 VORQSGDAAASVVVTGEGNEFAPVQNLQWSVSGTQVTLTWQAPASDKRTYVLYNESFTQ 240

QY 241 TLPNGWTMIDADGDNHNLSTINVTATHTGDMGAFSKSWTASGAKIDILSPDNYLYTP 300
 DB 241 TLPNGWTMIDADGDNHNLSTINVTATHTGDMGAFSKSWTASGAKIDILSPDNYLYTP 300

QY 301 KVTVPENKLSYVWSSQVPTNEHYGVFLSTTGNAAFTIKLLEETLGSOKPAPMNLVK 360
 DB 301 KVTVPENKLSYVWSSQVPTNEHYGVFLSTTGNAAFTIKLLEETLGSOKPAPMNLVK 360

QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDLDDVAVSGEGSSNDYTYV 420
 DB 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDLDDVAVSGEGSSNDYTYV 420

QY 421 YRDNVVIQNLAAATTFNQENVAPGQYNYCVKVTAGVSPKVKDVTVEGSNEFAHVQNL 480
 DB 421 YRDNVVIQNLAAATTFNQENVAPGQYNYCVKVTAGVSPKVKDVTVEGSNEFAHVQNL 480

QY 481 TGSAGVQKVTLLKWDAPN 497
 DB 481 TGSAGVQKVTLLKWDAPN 497

RESULT 2
 AA69483
 ID AA69483 standard; protein; 497 AA.
 AC AA69483;
 XX
 DT 22-DEC-1998 (first entry)
 DE
 DE Haemagglutinin protein hoga.
 KW Haemagglutinin protein; periodontal disease; vaccine; hoga.
 XX Porphyrinomas gingivalis.
 OS
 PN US5824791-A.
 PN
 XX 20-OCT-1998.
 XX
 XX 11-DEC-1995; 95US-00570311.
 XX
 XX 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX
 XX (UYEL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulsk-Fox A, Lepine G;
 PI WPI: 1998-582627/49.
 DR N-PSDB; AAV58870.
 DR
 XX

PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 protease poly:peptide(s)).
 XX
 PS Claim 1; Col 37-44; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 invention. This sequence represents the hoga haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 SQ Sequence 497 AA;

Query Match 100.0%; Score 2641; DB 2; Length 497;
 Best Local Similarity 100.0%; Pred. No. 2.5e-194;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLSLFLAVLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60
 DB 1 MRKLSLFLAVLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSVKDLRDPPIAGM 60

QY 61 ARIILEAHVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEYKVPVNA 120
 DB 61 ARIILEAHVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPFEYKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIYIVGEGVSKGNDYVVEAGKTYHFT 180

QY 181 VORQSGDAAASVVVTGEGNEFAPVQNLQWSVSGTQVTLTWQAPASDKRTYVLYNESFTQ 240
 DB 181 VORQSGDAAASVVVTGEGNEFAPVQNLQWSVSGTQVTLTWQAPASDKRTYVLYNESFTQ 240

QY 241 TLPNGWTMIDADGDNHNLSTINVTATHTGDMGAFSKSWTASGAKIDILSPDNYLYTP 300
 DB 241 TLPNGWTMIDADGDNHNLSTINVTATHTGDMGAFSKSWTASGAKIDILSPDNYLYTP 300

QY 301 KVTVPENKLSYVWSSQVPTNEHYGVFLSTTGNAAFTIKLLEETLGSOKPAPMNLVK 360
 DB 301 KVTVPENKLSYVWSSQVPTNEHYGVFLSTTGNAAFTIKLLEETLGSOKPAPMNLVK 360

QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDLDDVAVSGEGSSNDYTYV 420
 DB 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDLDDVAVSGEGSSNDYTYV 420

QY 421 YRDNVVIQNLAAATTFNQENVAPGQYNYCVKVTAGVSPKVKDVTVEGSNEFAHVQNL 480
 DB 421 YRDNVVIQNLAAATTFNQENVAPGQYNYCVKVTAGVSPKVKDVTVEGSNEFAHVQNL 480

QY 481 TGSAGVQKVTLLKWDAPN 497
 DB 481 TGSAGVQKVTLLKWDAPN 497

RESULT 3
 AA696030
 ID AA696030 standard; protein; 2628 AA.
 XX
 AC AA696030;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hoga.
 XX
 KW Haemagglutinin; hoga; periodontal disease; vaccine; antibody.
 OS Porphyromonas gingivalis; strain 381.
 XX
 XX Location/Qualifiers
 FT Key 5.21
 FT Peptide /label= Sig_peptide
 FT

XX WO9617936-A2.
 PN
 XX
 XX 13-JUN-1996.
 PD
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 XX 09-DEC-1994; 94US-00353485.
 PR
 XX
 XX (UYFL) UNIV FLORIDA.
 PA
 PA (UABR-) UAB RES FOUND.
 XX
 XX Progulskes-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX
 XX WPI; 1996-287181/29.
 DR
 DR N-PSDB; AAT30654.
 XX
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 6; Page 93-101; 153pp; English.
 PS
 XX
 XX P. gingivalis 381 haemagglutinin hgaA (AAR96030) was identified as the
 CC product of the hgaA gene (AAT30654) isolated as an EcoRV fragment of
 CC genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
 CC 24), can be obtd. from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
 CC also be used to detect the presence of anti-P. gingivalis antibodies and
 CC to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
 CC 2003 to standardise OS field)
 XX
 XX Sequence 2628 AA;
 SQ

Query Match 100.0%; Score 2641; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 2.9e-193;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKNSLFLSLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKIRTSKVKDLRDPVNA 60
 DB 1 MRKNSLFLSLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKIRTSKVKDLRDPVNA 60
 QY 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFVFANGTIPAGLYDPPFYKVPVNA 120
 DB 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFVFANGTIPAGLYDPPFYKVPVNA 120
 QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIVIGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIVIGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VQOQGGDAASVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVNLNFSFDTQ 240
 DB 181 VQOQGGDAASVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVNLNFSFDTQ 240
 QY 241 TLPNGWMTIDADGDGHNWLSLTINVTATHTGDMFSGKSWTASGGAKIDLSPDNVLTVP 300
 DB 241 TLPNGWMTIDADGDGHNWLSLTINVTATHTGDMFSGKSWTASGGAKIDLSPDNVLTVP 300
 QY 301 KVTVPNGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLLEETLGSCKPAPMNLVK 480
 DB 301 KVTVPNGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLLEETLGSCKPAPMNLVK 480
 QY 481 TGSVAGQKVTWKWDAPN 497
 DB 481 TGSVAGQKVTWKWDAPN 497

RESULT 4
 AAW69488
 ID AAW69488 standard; protein; 2628 AA.
 XX
 AC AAW69488;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hgaA.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN US5824791-A.
 XX
 PD 20-OCT-1998.
 XX
 PF 11-DEC-1995; 95US-00570311.
 XX
 PR 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX
 PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulskes-Fox A, Lepine G;
 XX
 XX WPI; 1998-582627/49.
 DR
 DR N-PSDB; AAY58875.
 XX
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease poly:peptide(s)).
 XX
 XX Claim 1; Col 91-110; 101pp; English.
 PS
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 XX Sequence 2628 AA;
 SQ

Query Match 100.0%; Score 2641; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 2.9e-193;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKNSLFLSLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKIRTSKVKDLRDPVNA 60
 DB 1 MRKNSLFLSLAVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKIRTSKVKDLRDPVNA 60
 QY 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFVFANGTIPAGLYDPPFYKVPVNA 120
 DB 61 ARTILEAHVDWEDGTGYQMLWDADHNOYGASIPESFVFANGTIPAGLYDPPFYKVPVNA 120
 QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIVIGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIIVIGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VQOQGGDAASVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVNLNFSFDTQ 240
 DB 181 VQOQGGDAASVVTGEGNEFAPVQNLQWSVSGQVTLTWQAPASDKRTYVNLNFSFDTQ 240
 QY 241 TLPNGWMTIDADGDGHNWLSLTINVTATHTGDMFSGKSWTASGGAKIDLSPDNVLTVP 300
 DB 241 TLPNGWMTIDADGDGHNWLSLTINVTATHTGDMFSGKSWTASGGAKIDLSPDNVLTVP 300
 QY 301 KVTVPNGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLLEETLGSCKPAPMNLVK 480
 DB 301 KVTVPNGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLLEETLGSCKPAPMNLVK 480

```

Db 301 KVVTFENGKLSYVSSQVPTWNEYGVFLSTGTGNEAANFIKLEETLGGDKAPMMLVX 360
Qy 361 SEGKLPAPYQERTIDLSAVAGQQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
Db 361 SEGKLPAPYQERTIDLSAVAGQQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
Qy 421 YRDNVVLAQNLAATFQENQENAVPGQYNYCVBKYTAGVSPKVKCDVTVGSGNEFAHYONL 480
Db 421 YRDNVVLAQNLAATFQENQENAVPGQYNYCVBKYTAGVSPKVKCDVTVGSGNEFAHYONL 480
Qy 481 TGSVAVGQKVKLWDAPN 497
Db 481 TGSVAVGQKVKLWDAPN 497

RESULT 5
AAW24786
ID AAW24786 standard; protein; 1706 AA.
XX
AC AAW24786;
XX
DT 17-OCT-2003 (revised)
DT 25-NOV-1997 (first entry)
XX
DE PrrR antigenic protein complex.
XX
KW Periodontal disease; cell surface protein; thiol protease; endopeptidase;
KW PrrR45; PrrR15; PrrR17; PrrR27; haemagglutinin; adhesin;
KW therapy; diagnosis; vaccine; antigen.
XX
OS Porphyromonas gingivalis; strain W50.
XX
PH Key
FT Peptide
FT 1. .227 Location/Qualifiers
FT /label= Pro-pro_peptide
FT /note= "44 kDa adhesin"
FT /label= PrrR45
FT /note= "45 kDa Arg-specific thiol protease"
FT /note= "720
FT /label= PrrR44
FT /note= "44 kDa adhesin"
FT /label= PrrR15
FT /note= "15 kDa adhesin"
FT /label= PrrR17
FT /note= "17 kDa adhesin"
FT /label= PrrR27
FT /note= "27 kDa adhesin"
XX
PN W09716542-A1.
XX
PD 09-MAY-1997.
XX
PF 30-OCT-1996; 96WO-AU000673.
XX
PR 30-OCT-1995; 95AU-00006275.
XX
PA (UWME ) UNIV MELBOURNE.
PA (VICT-) VICTORIAN DAIRY IND AUTHORITY.
XX
PI Reynolds EC, Bhogal PS, Slakeski N;
XX
DR WPI; 1997-272112/24.
DR N-PSDB; AAT78850.
XX

```

```

PT New antigenic protein complex from Porphyromonas gingivalis - comprising
PT Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
XX prevention and treatment of periodontal disease.
XX Example 1; Fig 8b; 68pp; English.
XX
CC A PrrR-PrrK cell surface protein of Porphyromonas ginivalis (PG)
CC comprises a 300 kDa complex composed a 45 kDa arginine-specific thiol
CC protease and 44, 15, 17 and 27 kDa adhesins encoded by the prrR gene
CC (AAT78850), and a 148 kDa lysine-specific thiol protease and 39, 15 and
CC 44 kDa adhesins (see AAW24787) encoded by the prrK gene (AAT78851). A
CC claimed antigenic complex comprises at least one multimeric protein
CC complex of PrrR and PrrK each containing at least one adhesive domain, the
CC complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
CC proteins of the PrrR-PrrK complex (see also AAW24780-85). It can be used
CC in a claimed composition to elicit an immune response directed against
CC PG, and in a claimed method of reducing the prospect of PG infection
CC and/or severity of disease. Antibodies directed against the complex are
CC claimed for use in treating PG infection. Unlike whole PG cells or other
CC previously prepared antigens based on fimbriae or the capsule, the PrrR-
CC PrrK complex or component parts are safe and effective antigens. (Updated
CC on 17-OCT-2003 to standardise OS field)
XX
XX Sequence 1706 AA;
SQ
Query Match 36.0%; Score 951.5; DB 2; Length 1706;
Best Local Similarity 31.9%; Pred. No. 1.3e-63; Indels 275; Gaps 20;
Matches 237; Conservative 74; Mismatches 156;
Qy 22 TAAAGQ-----GPKTAPSVTHQAVQK--GIR---TSKVXDLRDPDIPAGMARILAEH 68
Db 671 TATTQGGKVKLKWDA PSTKTNTATNTARSVDGIRLVLVSDAPELLRSQAEIVLEAH 730
Qy 69 DVWEDGTGYQMLWDADHNOYCASIPER--SFWFANGTIPAGLYDPFFPKYKPVNADAFSP 126
Db 731 DVWMDGSGYQILLDADHDQYGVIPSDTHLM--PNCSPANLPAFFPYTPENADPSCSP 789
Qy 127 TNFVLDGTASADIPAGTYDVYIINPNPGI-IYIVGEGVSGNDVVBAGTKYHTFVROG 185
Db 790 TNMIMDGTASVNPAGTYDFAIAPQANAKIWTAGQPTKEDDYVFEAGKYHFLMKMG 849
Qy 186 PGDAASVY----- 193
Db 850 SGDTELTISEGGSDYTYTVYRDGTIKKEGLTATTFEEDGVATGNEHYCVKTYAGVS 909
Qy 194 -----VTGEGNEFAPVQNLQWSVSCQTYLTWQAPAS-----DKRTYVINE 235
Db 910 PKVCKDVTVEGSNEFAVQNLGTSAVGQKVKLKWDA PNPNPNPNPNPNGTTLSE 969
Qy 236 SFDTQTLPNGWMTIDADGDGHNLSTINVTNTATHTGDGAMFSGKSWTASGAKIDLSPDN 295
Db 970 SFE-NGIPASWKITDADGDGHG--KPGNAPGIAGYNSNGCVYSEF-GLGGIGV-LTFDN 1025
Qy 296 YLVTPKVTVFENGKLSYVSSQ-VPTWNEHYGVFLSTGTGNEAANFTIKLEETL---GSD 351
Db 1026 YLITPDLPLNGGKLTFWVCAQDANYASEHYAVVAVSSTGNDASNFTNALLEETITAKGVR 1085
Qy 352 KPAPM----- 356
Db 1086 SPEAMRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRA DFTETF 1145
Qy 357 -----NLVKS----- 361
Db 1146 ESSTHGEAPAEWTIDADGGQGMCLSSGQLDLTAHGGTNNVVSFSGMNGMALNP DNYL 1205
Qy 362 ----- 361
Db 1206 ISKDVGTATKVKYVAVNCGFPGDHYAVMISKTGNAGDFTVVFETPNNGKGA RFL 1265
Qy 362 ----EGVKLPAPYQERTIDLSAVAGQQVYLAFRHFNSTGIFRLYLDV--AVSGGSSND 415
Db 1266 STEADGAKPOSVMERTVDLP--AGTK-YVAFRHYNSDLNILLDDDIQFTMGSPPTD 1322

```


CC modulate Arg-gingipain proteinase activity inhibitors
 XX Sequence 1704 AA;

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
 Best Local Similarity 31.9%; Pred. No. 4.4e-63;
 Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPDPAGMARILIEAH 68
 DB 671 TATTQCKVTLKWDAPSTKTNATTNTARSVDGIRLVLVSDAPELLRSQAIVLEAH 730
 QY 69 DVNEDGTGYQMLDADNQICASIPER--SFWFANGTIPAGLYDPFVKYKPVNADASFSP 126
 DB 731 DVNNDGSGYQILLDADHDQYGVTPSDTHILM--PNCVSPANLPAPFYTPENADPCSP 789
 QY 127 TNFVLDGTASADIPAGTYDVVIINPNFI--IYIVGEGVSKNDVYVPAQKTYHTVQROG 185
 DB 790 TNMIMDGTASVNPAGTYDEALAPQANAKIWIAGQGTKEDDYVFEAGKKYHFLMKMG 849
 QY 186 PGDAASV----- 193
 DB 850 SGBDTELTISGGGSDTYTYRDTGKIKEGLTATTFEEDGATGNHEYCVVKYTAGVS 909
 QY 194 -----VTGEGNEFAPVQNLQMSVSGTQVTLTWQAPAS-----DKRTVVLNESF 237
 DB 910 PKVKDVTVEGSEFAFVQNLGSAVCQKVTLLKWDAPNGTFNPNPNPFGTTLSESP 969
 QY 238 DTQTLNPGWTMIDADGCHNLSTINVINATHTGDGMFSKWTASGGAKIDLSPNYL 297
 DB 970 E-NGIPASWTIIDADGCHGW--KPGNAPGAGTNSGCVTSSESP-GLGGIGV-LTPDNYL 1025
 QY 298 VTPKVTVPENKLSYVSSO--VPMWNEHYGVFLSTTCNEAANFTIKLEETLGS----- 350
 DB 1026 ITPALDLPNGKLTFFWCAQDANAYSEHYAVYASSTGNDASNTNALLEETITAKGVRS 1085
 QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
 DB 1086 EAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLDEVEIKANGKRAFTTFES 1145
 QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
 DB 1146 STHGEAPAEWTTIIDADGGQGLCLSSGQLDLTAHGCTNNVASFSWNGMALNPDNYLIS 1205
 QY 386 ----- 385
 DB 1206 KDTGATKXYVAVNDGFFGDHYAVMISKTGNAGDFTVVFETPNNGKGARGLST 1265
 QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGGSSNDYTYTV 420
 DB 1266 EANGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDNLNYLLDDIQFTMGGSFTPTDYTYV 1325
 QY 421 YRDNVVIQNLATTFENQENAVAGQYNYCEVYKTAGVSKVDYTVESGNEFAHVQNL 480
 DB 1326 YRGTIKLKEGLTTFEEDGATGNHEYCEVYKTAGVSKVCNVNTI-NPTQFPNPKNL 1384
 QY 481 TCSAVGQKVTLKWDAPN 497
 DB 1385 KAQPDGDDVVLKWEAPS 1401

RESULT 9

AAU08938
 ID AAU08938 standard; protein; 1704 AA.

AC AAU08938;

XX 18-DEC-2001 (first entry)

XX P. gingivalis high molecular weight Arg-gingipain-2.

XX Periodontitis; antiinflammatory; Arg-gingipain-2; AG-2; immunogen.

XX

OS Porphyromonas gingivalis.

XX Key Location/Qualifiers
 FT Peptide 1..227
 FT /label= Prepro_peptide
 FT Region 599..619
 FT /note= "Region of homology with cysteine proteases"
 FT Region 670..674
 FT /label= Proteolytic_component
 FT Cleavage-site 719
 FT Protein 720..1091
 FT /label= HGP 44kDa
 FT /note= "Haemagglutinin protein component"
 FT Cleavage-site 1091
 FT Protein 1092..1429
 FT /label= HGP 17kDa
 FT /note= "Haemagglutinin protein component"
 FT Cleavage-site 1429
 FT Protein 1430..1704
 FT /label= HGP 27kDa
 FT /note= "Haemagglutinin protein component"

XX US6274718-B1.

XX 14-AUG-2001.

XX 25-JAN-2000; 2000US-00490931.

XX 10-SEP-1993; 93US-00119361.

XX 24-JUN-1994; 94US-00265441.

XX 09-SRP-1994; 94WO-US010283.

XX 08-NOV-1994; 94US-00336308.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Travis J. Potempa JS, Barr PJ, Pavloff N;

XX WPI: 2001-588904/66.

XX N-PSDB; AAS15242.

XX New recombinant DNA molecule which encodes high molecular weight (mature) Arg-gingipain protein, useful for immunization against inflammation and tissue damage, comprises enzymatically active protease component and hemagglutinin component.

XX Claim 1; Col 29-41; 56pp; English.

XX The invention relates to a recombinant DNA molecule encoding high molecular weight (mature) Arg-gingipain (AG) protein, which has an enzymatically active protease component (AG-2) and a haemagglutinin component, from P. gingivalis. The nucleic acid is useful for producing mature Arg-gingipain protein. Immunogenic compositions comprising Arg-gingipain are useful for immunising animals including humans against inflammatory response and tissue damage caused by an archaebacterium Porphyromonas gingivalis, which causes progressive periodontitis. Arg-gingipain is also useful for identifying agents that modulate Arg-gingipain proteinase activity, whether by acting on the proteinase itself or preventing the interaction of the proteinase with the protein in the gingival area, such as complement factors C3 or C5. The present sequence is Arg-gingipain-2

XX Sequence 1704 AA;

Query Match 35.8%; Score 944.5; DB 4; Length 1704;

Best Local Similarity 31.9%; Pred. No. 4.4e-63;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPDPAGMARILIEAH 68

DB 671 TATTQCKVTLKWDAPSTKTNATTNTARSVDGIRLVLVSDAPELLRSQAIVLEAH 730

QY 69 DVNEDGTGYQMLDADNQICASIPER--SFWFANGTIPAGLYDPFVKYKPVNADASFSP 126

Db 731 DVWNDGSGYQILLDADHDQYGVIPSDTHLW-PNCSSVPANLPAPFEYVTPENADPSCSP 789

Qy 127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKNDYVVEAGKYHTTVORQG 185

Db 790 TNMMDGTASVNIAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKHYFLMKMG 849

Qy 186 PGDAASVV----- 193

Db 850 SGDGTSLTISGGGSDYTYTVYRDGTJKIKEGLTATTFFEDGVATGNHCEYVVEKYTAGVS 909

Qy 194 -----VTGEGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----DKRTYVLNBSF 237

Db 910 PKVKCDVTVEGSENEFAPVQNLGSAVGQKVLTKWDAPNGTNPENPNPFGTTLSESF 969

Qy 238 DTQTLPGNWTMDADGDGHNWLSLTVNTATHTGCGAMFSGKSWTASGAKIDLSPDNYL 297

Db 970 E-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNNGCVYSESF-GLGGIGV-LTPDNYL 1025

Qy 298 VTPKVTVPENGKLSYVWSQ-VPWTHNEHYGVFLSTTGNAAANFTIKLEETLGS----- 350

Db 1026 ITPALDLPNGKLFVWCAQADANYASEHYAVYASSTGNDASNFTWALLEETITAKGVRS 1085

Qy 351 -----DKPAPMNLV-----KSEGVKL----- 366

Db 1086 EAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTTFES 1145

Qy 367 ----PAPYQERTID-----LSAYAGQV----- 385

Db 1146 STHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGNTVNVASFWSMGMALNPDNYLIS 1205

Qy 386 ----- 385

Db 1206 KDVGTATKVKYVAVNDGFGPDGHYAVMISKTGNAGDFTVVEETPNKNGKARFGLST 1265

Qy 386 -----YLAFRHNSGTIFRLYLDV--AVSGEGSSNDYTYV 420

Db 1266 EANGAKPOSVMERTVDLPAGTKYVAFRHYNGSLNYILLDDIQTMTGGSPTPTDYTYV 1325

Qy 421 YRDNVVIAONLAATTFNQENAPQYVVEKYTAGVSPKVKCDYVVEGSENEFAHVQNL 480

Db 1326 YRDGTJKIKEGLTETTFEDGVATGNHCEYVVEKYTAGVSPKCVNVTI-NPTQFNPVKL 1384

Qy 481 TGSAVGQKVLTKWDAPN 497

Db 1385 KAQPDGVDLKWEPAS 1401

RESULT 10

AAR96033

ID AAR96033 standard; protein; 1687 AA.

AC AAR96033;

XX

DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX

DE P. gingivalis haemagglutinin hage.

DE Haemagglutinin; hage; periodontal disease; vaccine; antibody.

XX

OS Porphyromonas gingivalis; strain FDC381.

XX

PN W09617936-A2.

XX

PD 13-JUN-1996.

XX

PF 11-DEC-1995; 95WO-US016108.

XX

PR 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX

PI Proguliske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX WPI; 1996-287181/29.

DR N-PSDB; AAT30656.

XX

PT Porphyromonas gingivalis genes and proteins - used in the detection and

PT vaccination against periodontal disease.

XX

PS Claim 5; Page 138-143; 153pp; English.

XX

CC P. gingivalis 381 haemagglutinin hage (AAR96033) was identified as the

CC product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.

CC The haemagglutinin can be obtd. from transformed host cells and used as a

CC vaccine to protect humans or animals against periodontal disease.

CC Expression in Salmonella cells allows prodn. of a live vaccine. The

CC haemagglutinin can also be used to detect the presence of anti-P.

CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic

CC appln. (Updated on 16-OCT-2003 to standardise OS field)

XX

SQ Sequence 1687 AA;

Query Match 35.7%; Score 943.5; DB 2; Length 1687;

Best Local Similarity 31.8%; Pred. No. 5.2e-63;

Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

Qy 22 TAAAG-----GPKTAPSVTHQAVQK-GIR---TSKVKDLRDPFAGMARILLLEAH '68

Db 652 TATTQGGKVLTKWDAPTKTNAITNTARSVDGRELVLSSDAPELLRSQAEIVLEAH 711

Qy 69 DVWEDGTGYQMLWDADHNOYGASIPEE--SFWFANGTIPAGLYDPFEPYKVPVADAFSP 126

Db 712 DVWNDGSGYQILLDADHDQYGVIPSDTHLW-PNCSSVPANLPAPFEYVTPENADPSCSP 770

Qy 127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKNDYVVEAGKYHTTVORQG 185

Db 771 TNMMDGTASVNIAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKHYFLMKMG 830

Qy 186 PGDAASVV----- 193

Db 831 SGDTEITISGEGSDYTYTVYRDGTJKIKEGLTATTFFEDGVATGNHCEYVVEKYTAGVS 890

Qy 194 -----VTGEGNEFAPVQNLQWSVSGQTVTLTWQAPAS-----DKRTYVLNE 235

Db 891 PKVKCDVTVEGSENEFAPVQNLGSAVGQKVLTKWDAPNGTNPENPNPFGTTLSE 950

Qy 236 SFTQTLPGNWTMDADGDGHNWLSLTVNTATHTGCGAMFSGKSWTASGAKIDLSPDN 295

Db 951 SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNNGCVYSESF-GLGGIGV-LTPDN 1006

Qy 296 YLVTVPKVTVPENGKLSYVWSQ-VPWTHNEHYGVFLSTTGNAAANFTIKLEETLGS----- 350

Db 1007 YLITPDLDPNGKLFVWCAQADANYASEHYAVYASSTGNDASNFTWALLEETITAKGV 1066

Qy 351 -----DKPAPMNLV-----KSEGVKL----- 366

Db 1067 SPEAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTTF 1126

Qy 367 ----PAPYQERTID-----LSAYAGQV----- 385

Db 1127 ESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGNTVNVASFWSMGMALNPDNYL 1186

Qy 386 ----- 385

Db 1187 ISKDVGTATKVKYVAVNDGFGPDGHYAVMISKTGNAGDFTVVEETPNKNGKARFGL 1246

Qy 386 -----YLAFRHNSGTIFRLYLDV--AVSGEGSSNDYTY 418

Db 1247 STEANGAKPOSVMERTVDLPAGTKYVAFRHYNGSLNYILLDDIQTMTGGSPTPTDYTY 1306

Qy 419 TYRDNVIVIAONLAATTFNQENAPQYVVEKYTAGVSPKVKCDYVVEGSENEFAHVQ 478

Db 1307 TYRDTGKIKKEGLTETTFEDGVATGNHCEYVVEKYTAGVSPKCVNVTI-NPTQFNPVK 1365

```

QY 479 NLGSAVGQKVTLLKWDAPN 497
Db 1366 NLKAQPDGDDVVLKWEAPS 1384

RESULT 11
AAW69495
ID AAW69495 standard; protein; 1687 AA.
AC
AC AAW69495;
XX
XX 22-DEC-1998 (first entry)
XX DE
XX Haemagglutinin protein hagE.
XX Haemagglutinin protein; periodontal disease; vaccine; hagE.
XX OS
XX Porphyromonas gingivalis.
XX PN
XX US5824791-A.
XX PD
XX 20-OCT-1998.
XX PF
XX 11-DEC-1995; 95US-00570311.
XX PR
XX 08-SEP-1988; 88US-00241640.
XX PR
XX 25-JAN-1991; 91US-00647119.
XX PR
XX 09-DEC-1994; 94US-00353485.
XX
XX (UYFL) UNIV FLORIDA.
XX PA
XX (UABR-) UAB RES FOUND.
XX
XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX WPI; 1998-582627/49.
XX DR
XX N-PSDB; AAW58681.
XX
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly(peptide(s)).
XX
XX Claim 1; Col 167-182; 101pp; English.
XX
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX CC
XX invention. This sequence represents the hagE haemagglutinin protein. The
XX CC
XX polypeptides are used to produce antibodies to organisms associated with
XX CC
XX periodontal disease. The antibodies are also used in purification and
XX CC
XX identification procedures. The genes and polypeptides are used as
XX CC
XX vaccines against periodontal disease
XX SQ
XX Sequence 1687 AA;

Query Match 35.7%; Score 943.5; DB 2; Length 1687;
Best Local Similarity 31.8%; Pred. No. 5.2e-63;
Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVQDLRDPDPAGMARIIIEAH 68
Db 652 TATTQGGKVTLLKWDAPSTKTNATNTARSVDGIRELVLLSVSDAPELLRSQAIVLEAH 711
QY 69 DVWEDGTGYQMLNDADHNOVCASIPER--SPWFANGTIPAGLYDPPEYKVPVNADASFSP 126
Db 712 DVWNGSGGYQILLADHDQGVIPSTHTLW-PNCSPVANLFAPEYTPVFNADPSCSP 770
QY 127 TNFVLDGTAGADIPAGTYDYVIINPNPGI-IYIVGEGVSGNDYVVEAGKTYHFTVQROG 185
Db 771 TNMIMGTASVNPAGTYDFAIAAPQANAKIWIAGQPTKREDDYVFEAGKKYHFLMKKMG 830
QY 186 PGDAASVV-----
Db 831 SGDGTETLISEGGSDYTYTVYRDGTIKI EKLATATTTFEEDGVATGNHEYCVKTYTAGVS 890
QY 194 -----VTGGGNEFAPVQNLQWSVSGQVTLLTWOAPAS-----DKRTYVLNE 235

```

RESULT 12

```

AAR96032
ID AAR96032 standard; protein; 1358 AA.
XX AC
XX AAR96032;
XX
XX 16-OCT-2003 (revised)
XX DT
XX 04-SEP-1996 (first entry)
XX
XX P. gingivalis hagD haemagglutinin.
XX DE
XX Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
XX KW
XX Porphyromonas gingivalis; strain FDC381.
XX OS
XX WO9617936-A2.
XX PN
XX 13-JUN-1996.
XX PD
XX
XX 11-DEC-1995; 95WO-US016108.
XX PF
XX 09-DEC-1994; 94US-00353485.
XX PR
XX (UYFL) UNIV FLORIDA.
XX PA
XX (UABR-) UAB RES FOUND.
XX
XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX WPI; 1996-287181/29.
XX DR
XX N-PSDB; AAT30655.
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
XX PT
XX vaccination against periodontal disease.
XX PS
XX Claim 5; Page 125-129; 153pp; English.
XX
XX P. gingivalis 381 haemagglutinin hagD (AAR96032) was identified as the
XX CC
XX product of the second open reading frame of the hagD gene (AAT30655)
XX CC

```

CC derived from P. gingivalis 318 genomic DNA. A first open reading frame
CC coded for hagD protease (see also AAR96031). The protease and
CC haemagglutinin can be obtd. from transformed host cells and used in
CC vaccines to protect humans or animals against periodontal disease.
CC Expression in Salmonella cells allows prodn. of live vaccines. The
CC haemagglutinin and protease can also be used to detect the presence of
CC anti-P. gingivalis antibodies and to raise monoclonal antibodies for
CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX SQ Sequence 1358 AA;

Query Match 30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%; Pred. No. 2.4e-53;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

QY 22 TAAAGGPKT---APSVTHQAVOKGIRTSKVKDLRD-----PIGAMARILEA 67
Db 328 TATTQGGKVTWKWDAPS-----AKAEASREVKKRIGDGLFVTIEPANDVRANEAKVILAA 382

QY 68 HDVMDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db 383 DNVGMDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

QY 127 TNFVLDGTASADIPAGTYDYVIINPNP--GLIYVGEV---VSKGNDYVVEAGKTYHTFTV 181
Db 442 QNIIVTQGEVVIPEGVYDICTINPEPASGKMWLAGDGNQPARYDDFTFEAGKKYFTFM 501

QY 182 QROGPGDAASVV----- 193
Db 502 RRAGMGDTMEVEDDSDPASVYTVYRDGTIKI EGLTATTFFEDGVAAGNHEVCVEVKYT 561

QY 194 -----VTGEGGNEFAPVQNLQWSVSGQVTLTWQAPAS-----DKRTYVLNE 235
Db 562 AGVSPKVKCDVTEGSENEFAPVQNLQWSVSGQVTLTWQAPAS-----DKRTYVLNE 235

QY 236 SFDQTQLPNGWTMIDADGHNWLTINVTATHTGCGAMPSKSWTASGAKIDLSPDN 295
Db 622 SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPDN 677

QY 296 YLVTPKVTPENGKLSYVWSQ-VPWTHNEHYGVFLSTTGNAAANFTIKLLEETLGS --- 350
Db 678 YLIITPALDLANGKLTWVCAQDANYASEHYAVYASSTGNDASFTNALLEETITAKGVR 737

QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 738 SPEAIRGIQTRQKTVDLPAGTKYVAFRHFQSDTMEFYIDLDEVEIKANGKRADFTET 797

QY 367 -----PAPYQERTID-----LSAVAGQOV----- 385
Db 798 ESSTHGEAPAEWTTIDADGDQDWLCLSSGQDLWLTAHGGTNVVASFSWNGMALNPNDYL 857

QY 386 ----- 385
Db 858 ISKDVGTATKYVYVNDGPFDDHYAVMISKTGTNAGDFTVVEETPNGINKGARGL 917

QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGESSNDYTY 418
Db 918 STEANGAKPQSWIERTYDLPAGTKYVAFRHYNCSDLLYILDDIQTFTMGSPPTDYTY 977

QY 419 TVYRNVVIAQNLAAATFENQENAVPGQYNYCEVKYTAGVSPKVKCDVTVGSGNEFAHQ 478
Db 978 TVYRDTKIKGLTETTFEEDGATGNHEVCVEVKYTAGVSPKVKCVNVTI-NPTQFNPK 1036

QY 479 NLTSAGVQKVTWKWDAPN 497
Db 1037 NLKAQPDGGDVVLKWEAPS 1055

RESULT 13
AAW69494
ID AAW69494 standard; protein; 1358 AA.
XX
AC AAW69494;

XX 22-DEC-1998 (first entry)
DT Haemagglutinin protein hagD.
DE Haemagglutinin protein; periodontal disease; vaccine; hagD.
KW Porphyromonas gingivalis.
OS US5824791-A.
XX 20-OCT-1998.
XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX WPI; 1998-582627/49.
XX N-PSDB; AAV58880.
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly:peptide(s)).
XX Claim 1; Col 145-158; 101pp; English.
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX invention. This sequence represents the hagD haemagglutinin protein. The
XX polypeptides are used to produce antibodies to organisms associated with
XX periodontal disease. The antibodies are also used in purification and
XX identification procedures. The genes and polypeptides are used as
XX vaccines against periodontal disease
XX SQ Sequence 1358 AA;

Query Match 30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%; Pred. No. 2.4e-53;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

QY 22 TAAAGGPKT---APSVTHQAVOKGIRTSKVKDLRD-----PIGAMARILEA 67
Db 328 TATTQGGKVTWKWDAPS-----AKAEASREVKKRIGDGLFVTIEPANDVRANEAKVILAA 382

QY 68 HDVMDGTGYQMLDADHNOYGASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db 383 DNVGMDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

QY 127 TNFVLDGTASADIPAGTYDYVIINPNP--GLIYVGEV---VSKGNDYVVEAGKTYHTFTV 181
Db 442 QNIIVTQGEVVIPEGVYDICTINPEPASGKMWLAGDGNQPARYDDFTFEAGKKYFTFM 501

QY 182 QROGPGDAASVV----- 193
Db 502 RRAGMGDTMEVEDDSDPASVYTVYRDGTIKI EGLTATTFFEDGVAAGNHEVCVEVKYT 561

QY 194 -----VTGEGGNEFAPVQNLQWSVSGQVTLTWQAPAS-----DKRTYVLNE 235
Db 562 AGVSPKVKCDVTEGSENEFAPVQNLQWSVSGQVTLTWQAPAS-----DKRTYVLNE 235

QY 236 SFDQTQLPNGWTMIDADGHNWLTINVTATHTGCGAMPSKSWTASGAKIDLSPDN 295
Db 622 SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPDN 677

QY 296 YLVTPKVTPENGKLSYVWSQ-VPWTHNEHYGVFLSTTGNAAANFTIKLLEETLGS --- 350
Db 678 YLIITPALDLANGKLTWVCAQDANYASEHYAVYASSTGNDASFTNALLEETITAKGVR 737

```

QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 738 SPEAIRIGIQTWKQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTET 797
QY 367 -----PAPQERTID-----LSYAGQOV----- 385
Db 798 ESSTHGEAPAEWTTIDADGDQDWLCLSSGQLDMLTAHGCTNVVAFSFWNGMALNPDNYL 857
QY 386 ----- 385
Db 858 ISKDVGTATKVKYKYAVNDGPFDDHYAVMLSKTGNAGDTVVFEETPNKGGARFGL 917
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTY 418
Db 918 STEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSLDLYILLDDIQFTMGSGSPTETDYY 977
QY 419 TVYRDNVVIAQNLAAATFNQENVAPGOYNCVEVKYTAGYSPKVKCKDVTVEGNEFAHVQ 478
Db 978 TVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGYSPKVCVNVVTI-NPTQFNPK 1036
QY 479 NLGTSVAVGQKVTWKWDAPN 497
Db 1037 NLKAQPDGSDVWLKWEAPS 1055

RESULT 14
AAR96029
ID AAR96029 standard; protein; 1732 AA.
XX
AC AAR96029;
XX
XX 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
XX P. gingivalis porphyain.
DE
XX Porphyain; haemagglutinin; periodontal disease; vaccine; antibody.
KW
XX Porphyromonas gingivalis; strain W12.
OS
XX
FH Key Location/Qualifiers
FT Region 688..708
FT /note= "Pro-Asn repeat region type 1"
FT Region 887..952
FT /note= "Pro-Asn repeat region type 2"
FT Region 946..967
FT /note= "Pro-Asn repeat region type 1"
FT Region 985..1006
FT /note= "Pro-Asn repeat region type 3"
FT Region 1041..1100
FT /note= "Pro-Asn repeat region type 4"
FT Region 1341..1405
FT /note= "Pro-Asn repeat region type 2"
FT Region 1430..1451
FT /note= "Pro-Asn repeat region type 3"
FT Region 1488..1547
FT /note= "Pro-Asn repeat region type 4"
FT Region 1607..1650
FT /note= "Pro-Asn repeat region type 2"
XX
XX WQ09617936-A2.
XX
XX 13-JUN-1996.
XX
XX 11-DEC-1995; 95WO-US016108.
XX
XX 09-DEC-1994; 94US-00353485.
XX
XX (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
XX Progulskse-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX

```

```

DR WPI; 1996-287181/29.
DR N-PSDB; AAT30653.
XX
PT Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
XX
XX Claim 5; Page 76-81; 153pp; English.
XX
CC P. gingivalis W12 cysteine protease, porphyain (AAR96029), was
CC identified as the product of the prp gene (AAR30653) isolated from P.
CC gingivalis W12 genomic DNA. The porphyain shows homology to the
CC haemagglutinins (see also AAR96026-28 and AAR96030-33) of P. gingivalis
CC 318. It can be obd. from transformed host cells and used as a vaccine to
CC protect humans or animals against periodontal disease. Expression in
CC Salmonella cells allows prodn. of a live vaccine. The porphyain and
CC haemagglutinins can also be used to detect the presence of anti-P.
CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
CC appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 1732 AA;
Query Match 30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 6.4e-53;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;
QY 22 TAAAGGGPKT---APSVTHQAVQKGIKRTSKVKDLRD-----PIPAGMARILEA 67
Db 693 TATTOQKQVTLKWEAPS-----AKKAGSREVIRIGDGLFVTIEPANDVRANEAKVVLAA 747
QY 68 HDVWEDGTGQMLMDADHNOYQASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db 748 DNVWGDNTGYQLLDADHNTFGSIVPATGFLF-TGTASSNLYSANFEYLVPANADPVVTT 806
QY 127 TNEVLDTASADIPAGTYDYIINPNP--GIIVVGEV---VSKGNDYVVEAGKTYHTFTV 181
Db 807 QNIIVTGGQEWVIPGGVYDYCIITNPEFASGKMVIAGDGGNQPARYDDTFEAGKKYKFTM 866
QY 182 QROGPGDAASV----- 193
Db 867 RRAGMGDGTMEVEDDPSASYTYTVYRDGTKIKEGLTATTFEEDGVAAGNHEVCVEKYT 926
QY 194 -----VTGEGNEFAPVQNLQWSVSGQVTVTLTQWAP-----ASDKRYVLN 234
Db 927 AGVSPKVKDQTVVEGSENEFAPVQNLATGSSVGQKVTLKWDAPNTPNPENPNPGTTL 986
QY 235 ESEDQTQTLNPGWTMIDADGHNWLTSTINVTATHTGDCAMPSKSWTASGGAKILDSPD 294
Db 987 ESPE-NGIPASWKTIDADGDGHGW-KPGNAPGIAGYNSNGCVYSEGF-GLGGIGV-LTPD 1042
QY 295 NYLVTPKVTVPENGKLSYVWSSQ-VPWTHYGVFLSTTGNAAANFTIKLEETLGS--- 350
Db 1043 NYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDSNTNALLETITAKGV 1102
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 1103 RSPKAIRIGIQTWKQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTET 1162
QY 367 -----PAPQERTID-----LSYAGQOV----- 385
Db 1163 FESSTHGEAPAEWTTIDADGDGQWLCSSGQLDMLTAHGGSNVVSSFSWNGMALNPDNY 1222
QY 386 ----- 385
Db 1223 LISKDVGTATKVKYKYAVNDGPFDDHYAVMLSKTGNAGDTVVFEETPNKGGARFG 1282
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTY 417
Db 1283 LSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSLDLYILLDDIQFTMGSGSPTETDYY 1342
QY 418 YTVYRDNVVIAQNLAAATFNQENVAPGOYNCVEVKYTAGYSPKVKCKDVTVEGNEFAHV 477
Db 1343 YTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGYSPKVCVNVVTI-NSTQFNVP 1401

```

QY 478 QNLGSAV--GQKVLKWDAP 496
 DB 1402 QNLTAEPNSMDALKKNAP 1422

RESULT 15
 ID AAW24787
 AC AAW24787;
 XX AAW24787;
 DT 17-OCT-2003 (revised)
 DT 25-NOV-1997 (first entry)
 XX
 DE PrtK antigenic protein complex.
 KW Periodontal disease; cell surface protein; thiol protease; endopeptidase;
 KW PrtK; PrtK48; PrtK39; PrtK15; PrtK44; haemagglutinin; adhesin; therapy;
 KW diagnosis; vaccine; antigen.
 OS Porphyromonas gingivalis; strain W50.
 XX

FH Key Location/Qualifiers
 FT Peptide 1..228
 FT /label= Pro-pro_peptide
 FT Cleavage-site 228..229
 FT Protein 229..737
 FT /label= PrtK48
 FT /note= "48 kDa Lys-specific thiol protease"
 FT Cleavage-site 737..738
 FT Protein 738..1156
 FT /label= PrtK39
 FT /note= "39 kDa adhesin"
 FT Cleavage-site 1156..1157
 FT Protein 1157..1291
 FT /label= PrtK15
 FT /note= "15 kDa adhesin"
 FT Cleavage-site 1291..1292
 FT Protein 1292..1732
 FT /label= PrtK44
 FT /note= "44 kDa adhesin"
 XX WO9716542-Al.
 XX 09-MAY-1997.
 XX 30-OCT-1996; 96WO-AU000673.
 XX 30-OCT-1995; 95AU-00006275.
 XX (UYME) UNIV MELBOURNE.
 XX (VICT-) VICTORIAN DAIRY IND AUTHORITY.
 XX Reynolds EC, Bhogal PS, Slakeski N;
 XX WPI; 1997-272112/24.
 XX N-PSDB; AAT78851.
 XX New antigenic protein complex from Porphyromonas gingivalis - comprising
 PT Arg- and Lys- specific thiol endo-peptidase(s), used in the detection,
 PT prevention and treatment of periodontal disease.
 XX Example 1; Fig 9b; 68pp; English.
 XX A PrtK-PrtK cell surface protein of Porphyromonas ginivalis (PG)
 CC comprises a 300 kDa complex composed a 48 kDa lysine-specific thiol
 CC protease and 39, 15 and 44 kDa adhesins encoded by the prtK gene
 CC (AAT78851), and a 45 kDa arginine-specific thiol protease and 44, 15, 17
 CC and 27 kDa adhesins (see AAW24786) encoded by the prtR gene (AAT78850). A
 CC claimed antigenic complex comprises at least one multimeric protein
 CC complex of PrtR and PrtK each containing at least one adhesin domain, the
 CC complex having a mol.wt. of over 200 kDa, and preferably comprises all 9
 CC proteins of the PrtR-PrtK complex (see also AAW24780-85). It can be used

CC in a claimed composition to elicit an immune response directed against
 CC PG, and in a claimed method of reducing the prospect of PG infection
 CC and/or severity of disease. Antibodies directed against the complex are
 CC claimed for use in treating PG infection. Unlike whole PG cells or other
 CC previously prepared antigens based on fimbriae or the capsule, the PrtK-
 CC PrtK complex or component parts are safe and effective antigens. (Updated
 CC on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1732 AA;

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
 Best Local Similarity 30.0%; Pred. No. 6.4e-53;
 Matches 222; Conservative 70; Mismatches 175; Indels 277; Gaps 22;

QY 22 TAAAGGPKT---APSVTHQAVQKGIKTSVKDLRD-----PIPMAGRIILEA 67
 DB 693 TATGQGVKTLKWEAPS-----AKAEGSRVRIKIGDGLFVTIEPANDVRANEAKVJLAA 747
 QY 68 HDVWEDGTGYQMLMDADHNOYGASIPESFWFANGTIPAGLYDP--FEYKVPVNADASFSP 126
 DB 748 DNVGCDNTGYQFLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLVPANADPVVVT 806
 QY 127 TNFVLDTASADIPAGTYDYVIINPNP--GIYIVGEG---VSKGNDYVVBAGTYHTV 181
 DB 807 QNLIIVTGGGVVIFGGVYDYCITNPEPASGRMWIAGDGGNQPARYDDFTFEAGKKYFTM 866
 QY 182 QROGEGDAASVV-----
 DB 867 KRAGMGDTMEVEDDSPAITYTYVRDGTIKIKEGLTATTFEEDGVAAGNHEYCEVEKYT 926
 QY 194 -----VTGEGNEFAPQNLQMSVSGQTVTLTWOAP-----ASDKRTYVIN 234
 DB 927 AGVSPKVKCDVTVEGSENEFAPQNLTGSSVGKVLKWDAPNGTNPNNPNPNTGTTLS 986
 QY 235 ESDFTQTLPNGWTMDADGCHNWLSTINVTATHTGCGAMFSKSWTASGGAKIDLSPD 294
 DB 987 ESFE-NGIPASNKTIDADGDGHW-KPGNAPGIAGYNSGCVYSBP-GLGGIGV-LTPD 1042
 QY 295 NYLVTPKVTPENGKLSYVWSQ-VPWTHYGVFSLTGTNEAANFTIKLLEETLGS--- 350
 DB 1043 NYLITPDLDPNGKLTFTWCAQDANYASEHVAVVASTGNDASNTWALLEETITAGV 1102
 QY 351 -----DKPAPNVLV-----KSEGVKL----- 366
 DB 1103 RSPKAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTET 1162
 QY 367 -----PAPYQERTID-----LSAVAGQV----- 385
 DB 1163 FESSTHGEAPAEWTTIDADGGQGMCLSSGQLDWLTAHGGSNVVSVSFWNGMALNPDY 1222
 QY 386 -----
 DB 1223 LISKDVTGATKVKYVAVNDGPPGDHYAVMISKGTNAGDFTVVFETPNKNGARFG 1282
 QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSCEGSSNDYT 417
 DB 1283 LSTEANGAKPOSWTIERTVDLPAGTKYVAFRHYNCSDLYILLDDIQFTMGSGPTTDTT 1342
 QY 418 YTVYRDNVVIAQNLAAATTFNOENVAPGOYNCVEVKYTAGVSPKVKCDVTVVEGSENEFAHV 477
 DB 1343 YTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCEVEVKYTAGVSPKVKCDVTVV-NSTQFPV 1401
 QY 478 QNLGSAV--GQKVLKWDAP 496
 DB 1402 QNLTAEPNSMDALKKNAP 1422

Search completed: May 18, 2004, 11:42:39
 Job time : 56.2572 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 18, 2004, 11:37:00 ; Search time 12.9765 seconds
(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-2
Perfect score: 2641
Sequence: 1 MRKLSLFSLAVALLSLLCWG.....QLTGSVAGQVKVTLKWDAPN 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------|--------------------|
| 1 | 2641 | 100.0 | 2628 | T28651 | hemagglutinin A - |
| 2 | 950.5 | 36.0 | 1526 | S49763 | gingipain R (EC 3. |
| 3 | 944.5 | 35.8 | 1704 | A55426 | gingipain R (EC 3. |
| 4 | 812.5 | 30.8 | 1732 | T30836 | lysine-specific cy |
| 5 | 494.5 | 18.7 | 991 | I40229 | arginyl endopeptid |
| 6 | 156 | 5.9 | 1231 | A48490 | endo-1,4-beta-xyla |
| 7 | 153 | 5.8 | 5291 | F90696 | hypothetical prote |
| 8 | 152 | 5.8 | 1234 | S72640 | endo-1,4-beta-xyla |
| 9 | 150 | 5.7 | 1483 | C37012 | probably celluloso |
| 10 | 147 | 5.6 | 1983 | G86643 | hypothetical prote |
| 11 | 144.5 | 5.5 | 781 | T36143 | probable secreted |
| 12 | 138.5 | 5.2 | 2468 | A83412 | hypothetical prote |
| 13 | 137.5 | 5.2 | 635 | S19011 | endo-1,4-beta-xyla |
| 14 | 137 | 5.2 | 2013 | A11489 | probable peptidogl |
| 15 | 135.5 | 5.1 | 1433 | A36734 | bacillopeptidase F |
| 16 | 135 | 5.1 | 1655 | T237835 | hypothetical prote |
| 17 | 132 | 5.0 | 1651 | JC1340 | outer membrane pro |
| 18 | 131.5 | 5.0 | 725 | A30255 | hypothetical prote |
| 19 | 129.5 | 4.9 | 1873 | T30947 | surface protein pr |
| 20 | 129 | 4.9 | 940 | AB1744 | internalin protein |
| 21 | 129 | 4.9 | 1349 | A11476 | cell surface prote |
| 22 | 128 | 4.8 | 1348 | A11115 | cell surface prote |
| 23 | 128 | 4.8 | 4936 | AH2515 | hypothetical prote |
| 24 | 127.5 | 4.8 | 5188 | T85547 | probable RTX fami |
| 25 | 127 | 4.8 | 2710 | A37052 | toxin A - Clostrid |
| 26 | 126.5 | 4.8 | 908 | A52254 | hypothetical prote |
| 27 | 126.5 | 4.8 | 1148 | T27635 | exo-poly-alpha-gal |
| 28 | 125.5 | 4.7 | 940 | A1374 | internalin protein |
| 29 | 125 | 4.7 | 1503 | T18266 | cycloinulo-oligosa |

| | | | | | | |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 124.5 | 4.7 | 868 | 2 | AF3204 | autotransporter pr |
| 31 | 124.5 | 4.7 | 875 | 2 | H90371 | proteinase import |
| 32 | 124 | 4.7 | 1616 | 2 | T17884 | S-layer protein - |
| 33 | 124 | 4.7 | 2013 | 2 | AD1129 | probable peptidogl |
| 34 | 123.5 | 4.7 | 208 | 2 | AF1718 | hypothetical prote |
| 35 | 122 | 4.6 | 729 | 2 | T35028 | probable glycosyl |
| 36 | 122 | 4.6 | 1217 | 2 | F97177 | alpha-glucosidase |
| 37 | 121.5 | 4.6 | 1530 | 2 | AH1396 | peptidoglycan anch |
| 38 | 121 | 4.6 | 1142 | 2 | C97080 | levanase/invertase |
| 39 | 120.5 | 4.6 | 587 | 2 | A55368 | transferred entry |
| 40 | 120.5 | 4.6 | 1448 | 2 | A12007 | Subtilase family p |
| 41 | 120 | 4.5 | 364 | 1 | JE0292 | fungai stress prot |
| 42 | 120 | 4.5 | 1939 | 2 | D97316 | probable S-layer p |
| 43 | 119 | 4.5 | 699 | 2 | A38368 | chitinase (EC 3.2. |
| 44 | 119 | 4.5 | 2205 | 2 | T08615 | aggregation factor |
| 45 | 118.5 | 4.5 | 694 | 2 | B84331 | hypothetical prote |

ALIGNMENTS

RESULT 1

T28651
hemagglutinin A - Porphyromonas gingivalis
C:Species: Porphyromonas gingivalis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C:Accession: T28651
R:Han, N.; Whitlock, J.; Proguliske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A:Title: The hemagglutinin gene A (haga) of Porphyromonas gingivalis 361 contains four :
A:Reference number: Z20494; MUID:97047672; PMID:8926061
A:Accession: T28651
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2628 <HAN>
A:Cross-references: EMBL:U01807; NID:g15524410; PID:g1469916; PIDN:AA017128.1
C:Genetics:
A:Gene: haga

| | | | | | | | |
|-----------------------|--------|---|----------|------------|---|--------|------|
| Query Match | 100.0% | Score | 2641 | DB | 2 | Length | 2628 |
| Best Local Similarity | 100.0% | Pred. No. | 2.1e-163 | | | | |
| Matches | 497 | Conservative | 0 | Mismatches | 0 | Indels | 0 |
| Gaps | 0 | | | | | | |
| Qy | 1 | MRKLSLFSLAVALLSLLCWGQTAAAGQPKTAPSVTHQAVQKGIKRTSKVKDLRDP | IPAGM | 60 | | | |
| Db | 1 | MRKLSLFSLAVALLSLLCWGQTAAAGQPKTAPSVTHQAVQKGIKRTSKVKDLRDP | IPAGM | 60 | | | |
| Qy | 61 | ARIILEAHVWEDGTGYQMLWDADHNOYGASIPESFWFANGTTPAGLYDPFEYKVPVNA | 120 | | | | |
| Db | 61 | ARIILEAHVWEDGTGYQMLWDADHNOYGASIPESFWFANGTTPAGLYDPFEYKVPVNA | 120 | | | | |
| Qy | 121 | DASFSPTNFVLDTASADIPAGTYDVIINPNPGIIVIVGEGVSKGNDYVVEAGKTYHFT | 180 | | | | |
| Db | 121 | DASFSPTNFVLDTASADIPAGTYDVIINPNPGIIVIVGEGVSKGNDYVVEAGKTYHFT | 180 | | | | |
| Qy | 181 | VQRCGPDAASVVVTGRCGNEFAFPVQNLQWSVSGTTLTQWAPASDKRTYVLNESDQTQ | 240 | | | | |
| Db | 181 | VQRCGPDAASVVVTGRCGNEFAFPVQNLQWSVSGTTLTQWAPASDKRTYVLNESDQTQ | 240 | | | | |
| Qy | 241 | TLPNGWTMIDAGDGHNLSTINVTATHTGDCAMFSKSWTASGGAKIDLSPNYLVT | 300 | | | | |
| Db | 241 | TLPNGWTMIDAGDGHNLSTINVTATHTGDCAMFSKSWTASGGAKIDLSPNYLVT | 300 | | | | |
| Qy | 301 | KVTVPENGLSYWSSQVPMTNHYGVFLSTGTGNEAANFTIKLLEETLGSQKPMNLVK | 360 | | | | |
| Db | 301 | KVTVPENGLSYWSSQVPMTNHYGVFLSTGTGNEAANFTIKLLEETLGSQKPMNLVK | 360 | | | | |
| Qy | 361 | SEGKLPAPQERTIDLSAYAGQVYLAFRHNSTGTFRILYLDVAVSGGSSNDYTYTV | 420 | | | | |
| Db | 361 | SEGKLPAPQERTIDLSAYAGQVYLAFRHNSTGTFRILYLDVAVSGGSSNDYTYTV | 420 | | | | |
| Qy | 421 | YRDNVVIAQNLAATFTFQENVAPGQYNYCIVEVKYTAGVSPKVKDVTVEGSNEFAHQNL | 480 | | | | |
| Db | 421 | YRDNVVIAQNLAATFTFQENVAPGQYNYCIVEVKYTAGVSPKVKDVTVEGSNEFAHQNL | 480 | | | | |

Db 421 YRDNVVIAQNLAATFNQENVAPGOYNYCEVKYTAGVSPKVKDVTVEGSEFAHQNL 480
 QY 481 TGSVAGQKVTLLKWDAPN 497
 Db 481 TGSVAGQKVTLLKWDAPN 497

RESULT 2
 S49763
 gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
 C:Species: Porphyromonas gingivalis
 C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997
 C:Accession: S49763
 R:Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
 submitted to the EMBL Data Library, November 1994
 A:Description: Cloning, sequence analysis and expression in Escherichia coli of prp1 of
 A:Reference number: S49763
 A:Accession: S49763
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1526 <ADU>
 A:Cross-references: EMBL:X82680
 C:Genetics:
 A:Gene: prp1
 C:Keywords: cysteine proteinase; hydrolase

Query Match 36.0%; Score 950.5; DB 2; Length 1526;
 Best Local Similarity 31.8%; Pred. No. 1.1e-53;
 Matches 236; Conservative 77; Mismatches 154; Indels 275; Gaps 20;

QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPAGMARIIIEAH 68
 Db 671 TATTGGKVTLLKWDAPSTKTNTATARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730

QY 69 DWEDGTGYQMLWDADHNDQYASIPER--SPWFANGTIPAGLYDPFEYKVPVNDASFP 126
 Db 731 DWNDSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPEYTVPENADPSCSP 789

QY 127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKNDYVVEAGKTYHFTVQROG 185
 Db 790 TNMMDGTASVNI PAGTYDYFAIAPOANAKIWIAGQGTKEDDYVFEAGKTYHFLMKMG 849

QY 186 PGDAASVV-----DKPA----- 193
 Db 850 SGDTGELTISRGGSDDYTYVYRDGDKIKEGLTATTPEEDGVAAAGNHEVCVEVKYTAGVS 909

QY 194 -----VTGEGNEFAPVQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 235
 Db 910 PKVCKDVTVEGSEFAFVQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 235

QY 236 SFDTQTLPGNWTMIDADGDGHNWLTINVTATHTGDMFSGKWTASGGAKIDLSPDN 295
 Db 970 SFE-NGIPASWKTIDADGDGHW-KGNAPAGTAGYNSNCVYSEF-GLGGIGV-LTPDN 1025

QY 296 YLTPKVTVPENGKLSYVSSQ-VPWTEHYGVFLSTTGNEAANFTIKLEETLGS----- 350
 Db 1026 YLITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASNTNALLEETITAKGVR 1085

QY 351 -----DKPA----- 354
 Db 1086 SPEAIRGIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADEFTTF 1145

QY 355 -----PMNLV 359
 Db 1146 ESSTHGEATAEWTTIDADGDGQNLCLSSGQLDLTAHGCTNVVSSFSWNGNALNPDNYL 1205

QY 360 ----- 359
 Db 1206 ISKDVGTGATKVKYVAVNDGPGDHVAVMISKTGTNAGDFTVVFETPGINKGARGFL 1265

QY 360 --KSEGVKLPAPYQERTIDLSAYAGQVYLAERHNSGTGIFLYLDDV--AVSGGSSND 415
 Db 1266 STEADGAKPQSVMIERTVDLP--AGTK-YVAFRHYNCSDLANIYLLDDIQFTMGSSPTFTD 1322

QY 416 YTYTVYRDNVVIAQNLAATFNQENVAPGOYNYCEVKYTAGVSPKVKDVTVEGSEFA 475
 Db 1323 YTYTVYRDGDKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVNVTV-NSTQFN 1381

QY 476 HVQNLTSAGVQKVTLLKWDAPN 497
 Db 1382 PVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 3
 A55426
 gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
 N:Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP;
 C:Species: Porphyromonas gingivalis
 C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
 C:Accession: A55426; D53113
 R:Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr,
 J. Biol. Chem. 270, 1007-1010, 1995
 A:Title: Molecular cloning and structural characterization of the Arg-gingipain protein
 A:Reference number: A55426; MUID:95138080; PMID:7836351
 A:Accession: A55426
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1704 <PAV>
 A:Cross-references: GB:U15282; NID:G557067; PIDN:AAA69539.1; PID:G557068
 R:Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
 J. Biol. Chem. 269, 406-411, 1994
 A:Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isola
 A:Reference number: A53113; MUID:94103245; PMID:8276827
 A:Accession: D53113
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 228-249 <PIK>
 A:Experimental source: H66
 A:Note: sequence extracted from NCBI backbone (NCBIP:141694)
 C:Keywords: cysteine proteinase; hydrolase

Query Match 35.8%; Score 944.5; DB 2; Length 1704;
 Best Local Similarity 31.9%; Pred. No. 3.2e-53;
 Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPAGMARIIIEAH 68
 Db 671 TATTGGKVTLLKWDAPSTKTNTATARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730

QY 69 DWEDGTGYQMLWDADHNDQYASIPER--SPWFANGTIPAGLYDPFEYKVPVNDASFP 126
 Db 731 DWNDSGYQILLDADHDQYGVIPSDTHTLW-PNCSVPANLFAPEYTVPENADPSCSP 789

QY 127 TNFVLDGTASADIPAGTYDYVIINPNPGI-IYIVGEGVSKNDYVVEAGKTYHFTVQROG 185
 Db 790 TNMMDGTASVNI PAGTYDYFAIAPOANAKIWIAGQGTKEDDYVFEAGKTYHFLMKMG 849

QY 186 PGDAASVV-----DKPA----- 193
 Db 850 SGDTGELTISRGGSDDYTYVYRDGDKIKEGLTATTPEEDGVATGNHEVCVEVKYTAGVS 909

QY 194 -----VTGEGNEFAPVQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 237
 Db 910 PKVCKDVTVEGSEFAFVQNLQWSVSGQVTLTWQAPAS-----DKRTYVINE 237

QY 238 DTQTLPGNWTMIDADGDGHNWLTINVTATHTGDMFSGKWTASGGAKIDLSPDNYL 297
 Db 970 E-NGIPASWKTIDADGDGHW-KGNAPAGTAGYNSNCVYSEF-GLGGIGV-LTPDN 1025

QY 298 VTPKVTVPENGKLSYVSSQ-VPWTEHYGVFLSTTGNEAANFTIKLEETLGS----- 350
 Db 1026 ITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASNTNALLEETITAKGVR 1085

QY 351 -----DKPAPMNLV-----KSEGVK----- 366
 Db 1086 BAIRGIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADEFTTFES 1145


```

QY 367 ----PAPYQERTID-----LSAVAGQOV-----385
Db 1146 STHGEAPAEWTTIDADGGGWLCLSSGQLDLTAHGTVNVASFWSNGMALNPDNYLIS 1205
QY 386 -----1385
Db 1206 KDVTGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFBEPTNGINKGGARFGLST 1265
QY 386 -----YLAFRHNSGICIFPLYLDDV--AVSGGSSNDYTYTV 420
Db 1266 EANGAKPQSVMIERTVDLPAGTKYVAPRHYNCSDNLNILLDDIQFTMGGSPTPTDYTYTV 1325
QY 421 YRNVVIAQLAATTFNQENAVPQYNYCQVVKYTAGVSKVCKDVTVGSGNEFAHVQNL 480
Db 1326 YRDGTKIKEGLTETTFEEDGVAIGNHEYCVVKYTAGVSKPCVNVTI-NPTQFPVKNL 1384
QY 481 TGSANGQKVLKWDAPN 497
Db 1385 KAQPDGGDVVLKWEAPS 1401

RESULT 4
T30836
lysine-specific cysteine proteinase porphyrain (EC 3.4.22.-) - Porphyrinomonas gingivalis
N:Alternate names: lysine-specific cysteine proteinase 1, 60K
C:Species: Porphyrinomonas gingivalis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30836; T30837; T30526; A53113
R:Barcoy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulski-Fox, A.; Lantz, J. Bacteriol. 178, 2734-2741, 1996
A:Title: Analysis of the prtp gene encoding porphyrain, a cysteine proteinase of Porphyrinomonas gingivalis
A:Reference number: Z20895; MUID:96213011; PMID:8631659
A:Accession: T30836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1732 <BAR>
A:Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R:Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z20896
A:Accession: T30837
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A:Cross-references: EMBL:U75366; NID:g2182811; PID:g2182812; PIDN:AAB60809.1
R:Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A:Title: IS195, an insertion sequence-like element associated with protease genes in Porphyromonas gingivalis
A:Reference number: Z20844; MUID:98298016; PMID:9632563
A:Accession: T30526
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A:Cross-references: EMBL:AF017059; NID:g2738802; PID:g2738803; PIDN:AAC26523.1
R:Pieke, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A:Title: Lysine- and arginine-specific proteinases from Porphyrinomonas gingivalis. Isolation and characterization of the genes
A:Reference number: A53113; MUID:94103245; PMID:8276827
A:Accession: A53113
A:Status: preliminary
A:Molecule type: protein
A:Residues: 229-249 <PIK>
A:Experimental source: H66
A:Note: sequence extracted from NCBI backbone (NCBI:141690)
C:Genetics:
A:Gene: prtp; prtk
C:Keywords: cysteine proteinase; hydrolase

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1.3e-44;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

```

```

QY 22 TAAAGGPKT----APSVTHQAVQKGIKTSKVKDLRD-----PIPAQMARIILEA 67
Db 693 TATTQCKVTLKWEAPS-----AKKAEGSREVKRIIGDLGVITIEPANDYRANEAKVVLAA 747
QY 68 HDVWEDGTGYQMLWDADHNOYASIPESFQWANGTIPAGLYDP--FEYKVPVNNADASFSP 126
Db 748 DNVWGDNTGYQFLLDADHNTFGSVIPATGFLP--TGTASSNLYSANFEYLVPANADPVVTT 806
QY 127 TNPVLGTSADIPAGTYDYVIINPNP--GLIYIVGEG--VSKGNDYVVEACKTVHFTV 181
Db 807 QNLIIYTGQGEVPIPGGVYDYCIITNPSPASGKMWIAGDGGNQPARYDDFTFEAGKYKTTFTM 866
QY 182 QROGPDGAASV-----193
Db 867 RRAGMGDGTMEVEDDPSASYTYTVYRDGPKIKEGLTATTFEEDGVAAGNHEYCVVKY 926
QY 194 -----VTGEGNEFAPVONLQWSVSGQVTVTWOAP-----ASKRYTVVLN 234
Db 927 AGVSPKVKCKDVTVGSGNEFAPVQNLATGSSVGQKVTILKWDAPNGTTPNPENPNPGTTL 986
QY 235 ESEDTOTLNGMTMIDADGHNWLSITINVTATHTGCGAMESKSWTASGGAKIDLSPD 294
Db 987 ESPE-NGIPASWKTIDADGGHGW-KPGNAPGATGNSGCVVYSESF-GLGGIGV-LTPD 1042
QY 295 NYLVTPKVTPVPEKGLSYVVSQ--VPWNEHGVFLSTTTCNEAANFTIKLEETLGS---350
Db 1043 NYLITPDLPLNGGKLTFWVCAODANYASEHYAVYASSIGNDASNTNALLEETITAKGV 1102
QY 351 -----DKPAPMLV-----KSEGVKL-----366
Db 1103 RSPKAIRGRIQTGWOKTVDLPAKTGYVAPRHFQSTDMFYDILDEVEIKANGKRAFDTE 1162
QY 367 -----PAPYQERTID-----LSAVAGQOV-----385
Db 1163 FESSTGEAPAEWTTIDADGGGWLCLSSGQLDLTAHGGSNVVSVFWSNGMALNPDNY 1222
QY 386 -----1385
Db 1223 LISKDVTGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFBEPTNGINKGGARFG 1282
QY 386 -----YLAFRHNSGICIFPLYLDDV--AVSGGSSNDYTYTV 417
Db 1283 LSTEANGAKPQSVMIERTVDLPAGTKYVAPRHYNCSDNLNILLDDIQFTMGGSPTPTDY 1342
QY 418 YTVYRNVVIAQLAATTFNQENAVPQYNYCQVVKYTAGVSKVCKDVTVGSGNEFAHV 477
Db 1343 YTVYRDTKIKEGLTETTFEEDGVAIGNHEYCVVKYTAGVSKPCVVDVTV-NSTQFPV 1401
QY 478 QNLITGSVAV--GQKVTILKWDAP 496
Db 1402 QNLITASQAPNSMDAILKKNAP 1422

RESULT 5
I40229
arginyl endopeptidase - Porphyrinomonas gingivalis
C:Species: Porphyrinomonas gingivalis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40229
R:Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A:Title: Structural characterization of arginogipain, a novel arginine-specific cysteine
A:Reference number: I40229; MUID:95168884; PMID:7864651
A:Accession: I40229
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-991 <RES>
A:Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match 18.7%; Score 494.5; DB 2; Length 991;
Best Local Similarity 25.5%; Pred. No. 2.9e-24;
Matches 124; Conservative 41; Mismatches 90; Indels 231; Gaps 7;

```

QY 22 TAAAG-----GPKTAPSVTHQVOK--GIR---TSKVXDLRDPDPAGMARILLEAH 68
 Db 671 TATTGQGVTLKWDAPSTKTATNTARSVDGIRELVLLSVSDAPELRSGQAEIVLEAH 730
 QY 69 DVNEDGTGYQMLWDADHNOYGASIPEE--SFWFANGTIPAGLYDPFFYKVPVNDASFSP 126
 Db 731 DVNEDGSYQILLDADHDQGVQVTPSDTHLM--PNCVPAFLPAPFEYTPVNDAPSCSP 789
 QY 127 TNFVLDGTASADIPAGYVYVVIINPNPGI--IYIVGEGVSGNDVVBAGTYHTVQROG 185
 Db 790 TNMIMDGTASVNIIPAGTYDFAIAAQAQANAKIWIAGQGPTEKDDYVFEAGKXHFLEKMG 849
 QY 186 PGDAASVVVTCGEGNEAPVQNLQMSVSGQVTLTWQAPASDKRTYVYNESFDTQTLPNG 245
 Db 850 SGGTELTIS----- 859
 QY 246 WTMIDADGDGHNWLTINVTNTATHTGDGAMFSKSWTASGAKIDLSPDNLVLPKVTVP 305
 Db 860 ----- 859
 QY 306 ENGLSVWSSQVPWNEHYGVFLSTTGNAAFTIKLEBTGLSDKPAFPAFLVKSQGVK 365
 Db 860 ----- 859
 QY 366 LPAPVQERTIDLSAVAGQVYLAFPHFNSTGIFRLYLDVAVSGEGSNDYTYTYRDNV 425
 Db 860 -----EGGSDYTYTYRDTG 875
 QY 426 VIAQNLAAFTTQENAVPQVNYCVKVTAGVSPKCVKVTBGSNEFAHVQNLTSQAV 485
 Db 876 KIKEGLTETTYRDAGMSAQSHYCVENKVAAGVSPKCVVDPDGADVTAQKPYTLTW 935
 QY 486 GQKVTIL 491
 Db 936 GKTITV 941

RESULT 6
 A48490
 endo-1,4-beta-xylanase (EC 3.2.1.8) A precursor - Thermoanaerobacterium saccharolyticum
 C:Species: Thermoanaerobacterium saccharolyticum
 C>Date: 03-May-1994 #sequence_revision 11-Apr-1997 #text_change 16-Jul-1999
 C:Accession: A48490
 R:Lee, Y.E.; Lowe, S.E.; Henrisat, B.; Zeikus, J.G.
 J. Bacteriol. 175, 5890-5898, 1993
 A:Title: Characterization of the active site and thermostability regions of endoxylanase
 A:Reference number: A48490; PMID:93388520; PMID:8376336
 A:Accession: A48490
 A:Molecule type: DNA
 A:Residues: 1-1231 <LE>
 A:Cross-references: GB:M97882; NID:G533366; PIDN:AAA21812.1; PID:G533367
 A:Experimental source: strain B6A-RI
 A>Note: mutation experiments support roles for Asp-537, Glu-600, and Asp-602; His-572 ca
 -495, Asn-539, and His-572 are supported by X-ray reference A55905 on another molecule
 C:Genetics:
 A:Gene: xynA
 C:Function:
 A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
 A:Pathway: xylan degradation
 C:Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
 A xylanase A cellulose-binding repeat homology
 C:Keywords: duplication; glycosidase; heat-stable protein; hydrolase; polysaccharide deg
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:34-1231/Product: endo-1,4-beta-xylanase A #status predicted <MAT>
 F:39-186/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1>
 F:196-340/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
 F:386-676/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
 F:683-854/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1>
 F:858-1043/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
 F:1056-1109/Domain: Thermotoga xylanase A cellulose-binding repeat homology <SLR1>
 F:1115-1168/Domain: S-layer repeat homology <SLR2>
 F:1179-1231/Domain: S-layer repeat homology #status atypical <SLR3>
 F:495,572,600,602/Active site: Glu, His, Glu, Asp #status predicted

F:537,539/Binding site: substrate (Asp, Asn) #status predicted

Query Match 5.9%; Score 156; DB 1; Length 1231;

Best Local Similarity 20.4%; Pred. No. 0.043;

Matches 92; Conservative 64; Mismatches 128; Indels 166; Gaps 23;

QY 106 AGLYDPEYKVPVNDASFPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSK 165

Db 22 ASLSP-----PIRVFADDNINLVNG---DFESGTDIGWIKQGNFTLAVTTEQAIG 72

QY 166 GNDYVVEAGTYHTVQROG-----GDAASV---VVTGB-GGNEFAPVQNLQ 209

Db 73 YSMKVTGRTQTY-----EGPAYSLFKMQGGSYSVSLKVLVSGQNSNPLITVTMFR 126

QY 210 WSVSGQTV-TLTQAPASD-----KRTYVLNESPD----- 238

Db 127 EDDNGRHYDVIWQKVSEDSWTVSGTYLDTIGTLKLYMYVESDPTLEYIIDVVV 186

QY 239 -----TQTLPNGWTMIDADGDGHNWLTIN-----VYNTATHTGAMFSKS 280

Db 187 TTQNPIQGVNVIANEITFENGNT-----SGWIGTSSVVKAVYGA-HSGDYSLLTIG 237

QY 281 WTAS-GGAKIDLSPDNLVTPKVTVPENGKLSYVWSSQVPTNEHYGVFELSTTGENAANF 339

Db 238 RTANWNGPSYDL-----TKIVPGQQYVDFVWK-----FVNGNDEQIKA 278

QY 340 TIKLEBTGLSDKPAF-----NLKSEGVKLPAPYQERTIDLSAVAGQVYLAFRHN 393

Db 279 TVKAT-----SDKONYIQVNDNFANVKEWTEIKGSP---TLPVADISGIYVESQ--N 328

QY 394 STGIFRLYLDVAVSGEGSNDYT-----YTVYRD----- 423

Db 329 PT--LEFYIDDFSVIGEISNNQITIQNDIPLYSVFKDYFPIGAVDPSRLNDADPHAQL 386

QY 424 -----NVVIAQNLAAFTTQENAVPQVNY 448

Db 387 TAKEFNMLVAEN-----AMKPESLQTEGNF 412

RESULT 7

F90696

hypothetical protein EC0542 [imported] - Escherichia coli (strain O157:H7, substrain R

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: F90696

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A39629; PMID:1156231; PMID:11258796

A:Accession: F90696

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-5291 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA033965.1; PID:G13360000; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: EC0542

Query Match 5.8%; Score 153; DB 2; Length 5291;

Best Local Similarity 24.3%; Pred. No. 0.52;

Matches 137; Conservative 70; Mismatches 202; Indels 154; Gaps 33;

QY 5 NSLFLAVLLSLLCWGQTAAAG--GPKTAPSVTHQVQKIGRTSKVKDLRDPDPAGMA- 61

Db 3303 NNGYTLTATVSDLAGNLGSKAGVTVDITAPVISENTVAGDDVINNVHEIQAIISGTAT 3362

QY 62 -----RIIEAHVYVEDGTGYQMLWDADHNOYGASIPFESF-WFANG--TIPAGLYDPFF 112

Db 3363 GAVAGDRLV-----VTIAGQQVTVTSTDSGN--WSVGVPASVISGLADGTVTISATID-- 3414

QY 113 FYKVPVNDASFPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVE 172

Db 3415 -----SAGNSSTQTHNVQNTAAVSLSVST-----ISGDNLIN-----AAE 3450
Qy 173 AKTYHFTVQPGPGDAASVVT-----GEGNEFAPVQ-NLQWSVSGQVTLTWQAPAD 227
Db 3451 AGSA--LTLSGTGTFATGTVVTVLLNGKYS--ATIQSGNSVSN--VPAADVAALSD 3503
Qy 228 KRTYVLNESPDITQTPNGMTMIDADGDGHNWLSSTINV-VNTA-----THTGDGAMFSK 279
Db 3504 GTSYTVSASQA-----DSAGNSSTQTHNVQNTAAVSLSVSTISGDNLIN-A- 3551
Qy 280 SWTASGAKIDLK--PDNYLWPKVTPVPEK--LSYVWSGQVPT-----NE 323
Db 3552 ---AEAGSALTSGTGTFATGTVVTVLLNGKYSATIQSGNSVSNVPAADVAALSDGT 3608
Qy 324 HYGVFLS---TTGNEAANFIKLEETLSGDKAP--MNLVSEGVKLPAPQERTIDL 377
Db 3609 SYTVSASQDSAGNSAT-----ASRSVAVDLTAPVISINTVSTD--RLNAEQQQPLTL 3661
Qy 378 ---SAYAGQV-----YLAFRHFNSTGIFRLYLDLDDVAVSGEGSSNDYTVTVRDNV 425
Db 3662 NGSTSAREVGTVTTFGGKTYTATVAANGTWALNPAVDLAALGQG----- 3707
Qy 426 VTAQNLAAATFNQENVAPGQYQYCVKVTAGVSPKCDVTVEGSEFAHVNQLTG--- 482
Db 3708 --AQTTTASV-NDRAGNPGQATHALTYD---TVAPT--TIATVAGDDIINNAEQLAGQTI 3760
Qy 483 ---SAVGOKVTL-----KWDA 495
Db 3761 SGTTTAEVGTGTVTFNGQWTA 3783
RESULT 8
S72640
end-1,4-beta-xylanase (EC 3.2.1.8) xynA precursor - Thermoanaerobacterium thermosulfurigenes
C;Species: Thermoanaerobacterium thermosulfurigenes
C;Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 16-Jul-1999
C;Accession: S72640; S72622
R;Matuschek, M.; Sahm, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes
A;Reference number: S72640
A;Accession: S72640
A;Molecule type: DNA
A;Residues: 1-1234
A;Cross-references: EMBL:U50952; NID:g1255235; PIDN:AAB08046.1; PID:g1255238
A;Experimental source: Strain EM1
R;Matuschek, M.; Sahm, K.; Zibat, A.; Bahl, H.
Mol. Gen. Genet. 252, 493-496, 1996
A;Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1
A;Reference number: S72621; MUID:97033555; PMID:8879252
A;Accession: S72622
A;Molecule type: DNA
A;Residues: 815-1234 <MAW>
A;Cross-references: EMBL:U50952
C;Genetics:
A;Gene: xynA
A;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C;Keywords: Glycosidase; hydrolase
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-1234/Product: endo-1,4-beta-xylanase XynA #status predicted <TXA1>
F;38-185/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1>
F;195-339/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2>
F;385-675/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SX1>
F;682-853/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1>
F;857-1042/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2>
F;1055-1108/Domain: S-layer repeat homology <SLR>
Query Match 5.8%; Score 152; DB 2; Length 1234;
Best Local Similarity 20.0%; Pred. No. 0.078;
Matches 97; Conservative 66; Mismatches 138; Indels 184; Gaps 23;
Qy 118 VNADASFPTNVLDTGTASADIPAGTYDVIINPNPGIIVYVGVSGKNDYVVEAGTKY 177

Db 28 IRAFADDTNINLSNG-----DFETGTDGWIQGNFTLEVTTEQAIGQYSMKVTGRQTY 83
Qy 178 HFTVQROGP-----GDAASV-----VVTGE--GGNEFAPVQNLQWSVSGQTV-TLT 220
Db 84 -----EGPAYSGFLGRMQKSGSYNSLVKRLVSGQNSNPFITVTFRBDNKGKHYDTIV 137
Qy 221 WQAPASDKRTYVLNESFDOTLPLNGMTMIDADGDGHNWLSSTINVNTATHTGDMAFSKS 280
Db 138 WQKQVSE-----DSWTVSGTYTLDYGT-----LKTLYMY----- 168
Qy 281 WTASGAKIDLSPDNYLWPKVTP-----ENGLKSYVWSGQVPTWNEHYGV--- 327
Db 169 -VESPDTELEYIDDDVVVTPQNPQVGNVTGTGTGNTSGWGTGSSVVKAVGVGAHS 227
Qy 328 ---FLSTTNEA-----ANFTIKLL-----EETLGSDEAP----- 355
Db 228 GGYSLTLTGRTANWNGPSYDLTGKIVPGQYVDFVWKFVNGNDTEQIKATVKTATSKON 287
Qy 356 -----MNLVSEGVKLPAPQERTIDLSAVAGQOVYLAFRHFNSTGIFRLYLDLDDVAVS 408
Db 288 YIQVNDPNNVNGEWEIKGSP---TLPVTDISGVSIYVESQ--NFT--LEFYIDDFSVI 340
Qy 409 GEGSSNDYT-----YTVYRD-----NVVIAQLAA 433
Db 341 GEISNNQIITQNDIPDLYSVFKQYFPPIGVAVDSSRLNDADPHAQLTAKHFNMLVAEN--- 397
Qy 434 TTFNQENVAPGQYQYCV-----VKYTAGVSPKCDVTVEGSEFAHVNQLTGSVAGQK 488
Db 398 -AMKPESLQTFEGNFTDNADKIVDY-----EIAHNNKMKRGH----- 433
Qy 489 VTILKW 493
Db 434 -TLLW 437
RESULT 9
C97012
probably cellulosomal scaffolding protein precursor, secreted, cellulose-binding and coi
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97012
R;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1483 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78886.1; PID:g15023809; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0910
Query Match 5.7%; Score 150; DB 2; Length 1483;
Best Local Similarity 20.2%; Pred. No. 0.14;
Matches 127; Conservative 81; Mismatches 213; Indels 208; Gaps 29;
Qy 27 GGPKTAP-----SVTHQAVQKGIKTSKVKDLRDPDIPAGMARILEAHHDVWEDGTGY- 77
Db 522 GTPVWTFPSQINVEGSGATDQPVKIDLNGTLKDWQDQ-----SGKTLVQGTDTVTDTGIT 577
Qy 78 -----QMLWDADHNOYGAS-----IPESFWFANGTIPAGLYDPEYKVPV 118
Db 578 LSQSYLAGIAGQYTLTLDNFNGGASQITITINVVKNETVKLSVGTSGNPGD---TVKVPV 635
Qy 119 N-----ADASFSTNF-VLDGTASADIPAGT---YDVVINPNPGIIVY--- 159
Db 636 TISQVSTPUGLICMWDISYDASKFTVKDVLPTDLVKDNTDYSFVNTSTFGKLSITDTP 695
Qy 160 -----GEGVSKGNDYVVEAGTKY---YHFTVQROGFGDAASVVVTGEGNEFAPVQNL 208

Db 696 TLNYPISVGLAYLDPIINSNATAGDSALT---DPATLIVADEND-----KDI 743
 QY 209 QWSVSGTTLTWOAPASDKRYTVLNSPDT--QTLP-----NGWTMIDA-DGDGH 256
 Db 744 KDAASNGKITVTSAPV--VQSSVVNTSSVTYDQNAFDQAVSITFNGNTVKDVKDASG- 800
 QY 257 NWLSTINVTATHTGGAMFESKSWTASGG--KIDLSPDN-----YLVTPKVTUPE 306
 Db 801 ---NTLKAGSDYATSDGILTSQSYLATLAAGTYYTIDFSAGNAGTFTVVVKGKTUVGS 857
 QY 307 NGKLSYVSS-----QVPWT---NEHYGVFLSTTGNAAFTIK-----LLEET- 347
 Db 858 ATTLAVGTVSGKAGTVKVPVITISKVTPVGLICAEIDYDASKFTVKDVLPTDLVKDTD 917
 QY 348 ---LGSKDPAPNL-----VKSQVGLPAPYQERTIDLSAYAGQQVYLAFRH 391
 Db 918 NYSFIVNTSTPGKISITFTDPTLANYPISADGI---LAYLDFIINSNATAGDSALT--- 971
 QY 392 FNSGTIFRLYLDVAVSGEGSSNDYT----- 417
 Db 972 -NPSGFIADENDKDIQDAASNGKITVGTSTPVAENSVNTSSVTYDQNAFDQAVSITL 1030
 QY 418 -----YTVRDNVIAQNLAAFTTQENVAPQYNYCEVRY-T 455
 Db 1031 NGNTITDVKDASGNTLKAGSDYTVTSDGILTQSGLAT-----LAAGTYTIVDFSGAGN 1084
 QY 456 AGVSPKVKCVTVVSGNEFAHVQNLTGSA 484
 Db 1085 AGTFTVVVKAKTVVSSATTLAVGTVSGKA 1113

RESULT 10

G86643
 Hypothetical protein ybeF [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis se
 A:Reference number: A86625; MUID:21235186; PMID:11337471
 C:Accession: G86643
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1983 <STO>
 A:Cross-references: GB:A8005176; PID:g12723000; PIDN:AAK04249.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: ybeF

Query Match 5.6%; Score 147; DB 2; Length 1983;
 Best Local Similarity 21.7%; Pred. No. 0.32;
 Matches 117; Conservative 77; Mismatches 181; Indels 164; Gaps 30;
 QY 39 AVQKGIATSKVLDLPIAGMARIILEAHWDWEDGTGYQLMDADHNYQASIPESFW 98
 Db 628 ATQGI-----VKNIDGDI-----DYENREIPWKIDINSPG-----YW 660
 QY 99 FANGTIPAGLYDPFYPKVPVNDASFPNPF-VLQGTASADIPAGTYDYVINPNPGLIY 157
 Db 661 MENWSL-----EDKMSGLFLENTFOIDKTAGNKVLSP-ETYLKTTAGFSV 709
 QY 158 IVGEGVSKGND--YVVEAGKTYHTVQROGPG-----DAASVVTGEGN----- 200
 Db 710 SENSPLKEGTHKQIKYKTKFDTSDVINGSGHEGDIKFNVDASTWKDKNQGDHNDH 769
 QY 201 -EPAPQNLQWSVSGTTLTWOAPASDKRYTVLNSPDTQTLPLNGWTMIDAGDGHNL 259
 Db 770 KEFKPIPPQY--NGQK-SGSYNA-TSKKITWTIAANFNQELNSA-SITDPISDQNV 824
 QY 260 S-----TINYNT-----AHTGGMFESKSWTASGAKIDLSPDN-----YLVTPKVT 303

Db 825 SGSAKVEATINKNGTYTLGAETSD--MGKIVESKGVKVEL-PDGTAKAYVLIFETS 881
 QY 304 VP-----ENGKLSYVSSQVPTNHEHGVFLSTTGN----- 335
 Db 882 LEGNLINQEKYKKAFTTKNDISHLSASV--TPAHQGEFTKDGSSQSDTSNYYVWKLT 939
 QY 336 --AANFTIKLEETLGSKDPAPMNLVKSEGVKLPAPOERTIDLSAYAGQQVYLAFRHN 393
 Db 940 VNASQSTLKNVEYT---DNPSNQYIVAKDILI---YGTSIDASG-----N 979
 QY 394 STGIFRLYLDVAVSGEGSSNDYTYYVNDNVIAQNLAAFTTQENVAPQYNYCEVX 453
 Db 980 ITENKNILIE-----QSGKYSVDIOTDNSTGAQTKIKPLSEINTA-----YVVE-- 1024
 QY 454 YTAGVSPKVKCVT---VEGSNE-----FAHVQNLTGSAVQK--VTLKWDAP 496
 Db 1025 YRALITSDKANDVVTQAHITGNEKIEQDVKDPVTVNHSANGSKSVLEKVG 1083
 RESULT 11
 T36143
 Probable secreted proteinase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2003
 C:Accession: T36143
 R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A:Reference number: Z21598
 A:Accession: T36143
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-781 <SEE>
 A:Cross-references: EMBL:AL096852; PIDN:CAB51001.1; GSPDB:GN00070; SCOEDB:SCE19A.20c
 A:Experimental source: strain A3(2)
 C:Genetics:
 C:Superfamily: secreted virulence protease, Inha type

Query Match 5.5%; Score 144.5; DB 2; Length 781;
 Best Local Similarity 21.2%; Pred. No. 0.13;
 Matches 111; Conservative 58; Mismatches 164; Indels 191; Gaps 31;
 QY 24 AAQGGPKTAP-SVTHQAVQGIKTSKYKLDLDPDPAG-----MARIILEAHQVWE----DG 74
 Db 205 ARYGSNKCDPNCWAYAVQDGV-TAWIADQK---AAGRTDAQIKSOLTQYDQWDRYDFD 260
 QY 75 TG-----YQMLWDADHNYQASIPESF-----WFGNTIPAGLYDPPEYKVPV 118
 Db 261 DGFNEPDGYIDHFQIVHAGDESGAGGAGQEDAIWAHRWYAFGT-DAGATGPADNK--- 316
 QY 119 NADASFSTNFVLDTASADIPAGTYDYVINPNPGLI-IYIVGEGVSKGNDYVVEAGKTY 177
 Db 317 -----LGQTQIGDTGIWVDYTIQPENGGLGV-----AHEYHDLGLPD 356
 QY 178 HFTVQROGPGDAASVVTGEGNEFAPVQNLQWSV--SGQTVTLTWOAPASDKRYTVLNE 235
 Db 357 HY-----DTAG-----GDNSTGF-----WTLMSSG-----SWLGTGR-----NE 385
 QY 236 SFDTQTLPGNWTMIDADGDGHNLSTINVTATHTGGMFESKSWTASG----- 285
 Db 386 IGDLPDGMNDKIQ-----GWLN-----YDTA-----KAGVNSHKLGLAEYNTKHQ 430
 QY 286 GAKIDLSPDNLVTPKVTVPENGKLSYVSSQVPTNHEHGVFLSTTGNAAFTI----- 341
 Db 431 GLAVELPKK--VTTEIVTPAEGTQWMSGNDLKNL-TSRVDLTGKSAASLTLDGWY 487
 QY 342 -----KLEETL-----GSKDPAPMNLVKSEGVKLPAPOE 372
 Db 488 DTEADYDFLYTEVSTGDNWTAIDGTFDGNPIQRDGSXKPSALSATVDAYG-KLVYP--- 543
 QY 373 RTIDLSAYAGQQVYLAFRHNSTGIFER--LVLDVAVSGEGSSNDYTYYVNDNV----- 425
 Db 544 -----LDAYAGKKIDURFRIQDTGGLAMKGFATAIATADGE-----TLFSDNAETADD 593

QY 426 -----VIAQNLAAATTENQENVAPQYNY 448
Db 594 AWTAVGTRKGASFTKEYAQYIAENRQVSYD-KILKTPINF 636
RESULT 12
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83412
R:Stover, C.K.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
adman, S.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; MID:g9947856; PIDN:AAG05263.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874
Query Match 5.2%; Score 138.5; DB 2; Length 2468;
Best Local Similarity 23.74; Pred.No. 1.6;
Matches 107; Conservative 47; Mismatches 166; Indels 131; Gaps 23;
QY 103 TIPAGLYDPPEXKVPVNNADASFPTNEFLDGTASADIPAGTYDVIINPFGIYIVGEG 162
Db 319 TDPAG-----NNSTPVTVEAP-----DTTA-----PAPATD-----VQVAPDG 351
QY 163 VS-KGNDYVVEAGKTHFTVQROGPDAAVSVVTGEGNEFAPVQNLQNSVGGTQVTLTW 221
Db 352 SSVTGN---AEPGAT--VGVDTDGQPDVTVVVVGPGSGFEVPLN--PPLTNGETVTIV 404
QY 222 QAPASDKRTVVLNESFTQILPN-----GMTMDADGDG----- 256
Db 405 TDPAGNSSTPVTAEAPDPAPQVNASGSLGTABAGTVITDGNPNIGQTSADAN 464
QY 257 -NMLST-----INVT-----NTATH--TGEGAM-----FSKSWTASG 285
Db 465 GNSWFTPGSQLPDGTVVNVVARDAGNSPATSTITDGVAPNAPVVPFSGSELSTGTAEP 524
QY 286 GAKIDLSPDNYLTPKVTVPENGKLSVWSSQVPTWNEHYGVFLSTTGNENAAFTIKLE 345
Db 525 GSSVTLTDGNGNPIGQTTADANGWSFTPTPLP-----DGTVVNVVARDAGNSPPAS 579
QY 346 ETLSGDKPAPNLVKSEGVKLPAFYQERTIDLSAYAGQVYLAFFHFNSTGIFRLYLDVV 405
Db 580 VTVDVAPATPTVDPSNGT-----TSLGTAEPSGSSVTLTDGNPNIG----- 621
QY 406 AVSGEGSSNDYTF-----VYRDNVVIAQNLAAATTFQENVAPQYNYCVEYKTAGVSPK 461
Db 622 QVTADGSGN-WTFPTSPPLNGTVV---NATATDPSGNASSPAS-----VTVDAVAPATPV 673
QY 462 V--CKDVTGSGNEFAHVQNLIT-----GSAVGQ 487
Db 674 VNPENGTTLSTAGTAPFGATVTLTDGNPNIGQ 704
RESULT 13
S19011
endo-1,4-beta-xylanase (EC 3.2.1.8) - Bacillus polymyxa
C:Species: Bacillus polymyxa
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 15-Oct-1999
C:Accession: S19011
R:Gosabes, M.J.; Perez-Gonzalez, J.A.; Gonzalez, R.; Navarro, A.
J. Bacteriol. 173, 7705-7710, 1991
A:Title: Two beta-glycanase genes are clustered in Bacillus polymyxa: molecular cloning,

A:Reference number: S19011; MUID:92041687; PMID:1938968

A:Accession: S19011

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-635 <GOS>

A:Cross-references: EMBL:X57094; NID:g48815; PIDN:CAA40378.1; PID:g48816

A>Note: the authors translated the codon GAA for residue 78 as Gly, CCT for residue 272

C:Function:

A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans

A:Pathway: xylan degradation

C:Superfamily: Clostridium xylanase A repeat homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:408-502/Domain: Clostridium xylanase A repeat homology <CXA>

Query Match 5.2%; Score 137.5; DB 2; Length 635;

Best Local Similarity 21.2%; Pred.No. 0.27;

Matches 123; Conservative 65; Mismatches 176; Indels 217; Gaps 32;

QY 22 TAAAGGPKTAP-----SVTHQAVQKGIKTSKVKDLRDPAGMARILLEAHDYWEDGTGY 77

Db 156 TADTPIGPWTDLGKALVTHST-----PGMAGVTWLFDPVAVLDDGTGY 200

QY 78 QMLWDADHNOYQASIPESFWFANGIPAGLYDPPEYKV-----PNADA 122

Db 201 --LYS-----GGGIPNES-----DPASIANPKTARVILKLGADMTSVIGSATTIDAPY 245

QY 123 SESPT-----NFVLDGTASADIPAGTYDVVII-NP-----NPG 154

Db 246 LFEDSGIHKNYKYYSYCINFA--GTHFQOYPAGEIGYMSVNDPMGPFTYKGFHFNKY 303

QY 155 IYIVG-----EGVSKGNDYVVEAGKTY---HFT-VQRQGPQDAA 190

Db 304 TFEVGSGNNHHAFFENFKNWYVVVHAQTVSKAQ---IGAGKGYRSPHINKLVHKEDGSIS 360

QY 191 SVV--VTGEGNEFAPVQNLQNSVGGTQVTLTQWAPASDKRTVVLNESFTQILPNCWTM 248

Db 361 EVQGNMTG-----IAQLSNMNPYTRVEAETIAQAGVTTEPT----- 397

QY 249 IDADGDGHNWLSTINVTNTATHTGDGAMFSKWTASGAK-----IDLSP 293

Db 398 -QASGGP-----ISNLNTN--IHNGDIWAVKADFGSAGAKTFKANVTATNVGGNIEVLDS 451

QY 294 DNYLVTPKVTVPENGKLSYV--VSSQVPTWNEHYGVFLSTTGNAAAN-----FTIKL 343

Db 452 ETGPELVGSLKVPSTGGMQWREVEETTINNATGVHNYLVFTGSGGNLLDLDAWQFTPT 511

QY 344 LETLSGDKPAPNLVKSEGVKLPAFYQERTIDLSAYAGQVYLAFFHF--NSTGIFRLY 401

Db 512 GGNITIKVEAENMKIGTYAGKISAPFD-----GVALLYANAD-VYSYSQYFANST----- 560

QY 402 LDDVAVSGEGSSNDYTVTVYRDNVVIAQNLAAATTFQENVAPQYNYCVEYKTAGVSPK 461

Db 561 -HNLISVRG-ASSNAGTAKV---DLVIG-----GTVGGSFNF-----TGKTP 597

QY 462 VKDVTGSGNEFAHVQNLITGSAVGQKVTIL-----KWDA 495

Db 598 V-----QILSNITHATGDOEIKALTSDDGTWDA 626

RESULT 14

A11489

probable peptidoglycan bound protein (LPXTG motif) lin0457 [imported] - Listeria innocua

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: A11489

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Feihl, H.

D.; Jones, L.N.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: A11489
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2013 <GLA>
 A;Cross-references: GB:ALU592022; PIDN:CA95689.1; PID:gl6412898; GSPDB:GN00178
 A;Experimental source: strain Clip11262
 C;Genetics:
 A;Gene: lin0457

Query Match 5.2%; Score 137; DB 2; Length 2013;
 Best Local Similarity 20.9%; Pred. No. 1.5; Indels 144; Gaps 30;
 Matches 103; Conservative 82; Mismatches 163; Indels 144; Gaps 30;
 Qy 74 GTGQCM-LWDADHNGYASIEES---FWFANGTIPAGLYDPFYPKVPVNADASFSTNF 129
 Db 1185 GDYVEIALFDKDGKEVQSAIKTNKGEFSFD---VAIKNPADFKLVKTPAP---TRINF 1237
 Qy 130 VLDGTASADIP---AGTYDYVVIINPQIIVIVGEGSKGNDYVVEAGK-TYHTTVORQG 185
 Db 1238 VY---SAKNELFMSTKEYTINSVPGV-----GGVAE--IYITETSKPTTKILDKAV 1286
 Qy 186 PGDAASVVVTGEGNEFAPVONLQWSV-----SGQTV-----TLTW 221
 Db 1287 TPNA-----ITIESSDEAREVTN-EMTVVDSNGTVVSGTGNIRIPNDEGYIAKNTATD 1341
 Qy 222 QA--PASDKRYVVLNESDPTDTLPNGWTMIDAGDGH-----NWLSTINVTATHTGD 273
 Db 1342 EAGNTASDEKTFDIDYTVPTLVNQ-----DASAEVNSTEANDWIKPLNVAATDTHGN 1396
 Qy 274 ---GAMFSK-SW-----TASGAK-----IDLSPDNVLTVPKVTVPBNG 308
 Db 1397 ITPVVDYSGKWDVLGTPYTVTTVTDASGNKATQTNLRIVDTTSPITLIT-----NN 1449
 Qy 309 KLSYVWSVQVPTNBYHYGFLSTTCNE-----AAEFIKLLEETLG 349
 Db 1450 PLTVSIENWRKLEBELYKAAGLIGDNYDLAPGQSVQNPQPMVFTSNFS--TIFSDIA 1507
 Qy 350 SDPK----APMNLVSEGVK-LPAPYQERTIDLSAYAGQVYVLAFRHFNSTGIFRLYLD 404
 Db 1508 SVKPGQYQVQVNLADSSGQALPQITINVTWDMGPVVKADNVSY-HVNTKTEAEFPQD 1566
 Qy 405 VAVSGEGSSNDYTYVYRDNVVIAONLAATTFNQNV---APGQYNYCEVYKTAGVSPK 461
 Db 1567 ARLDVTDNNDT-----NLIITSNFA-----EKNVLNPKGYEVTISATDTKG--NQ 1612
 Qy 462 VKCDVTVEGSNE 473
 Db 1613 TTKEITVQVSKD 1624
 RESULT 15
 A36734
 bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
 C;Species: Bacillus subtilis
 C;Date: 19-May-2000 #sequence revision 19-May-2000 #text change 16-Jun-2000
 C;Accession: A36734; A35131; A35750; B35750; S08223; JN0335; I39849; B69596; J00084
 R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
 J. Bacteriol. 172, 5520-5521, 1990
 A;Reference number: A36734; MUID:90368623; PMID:2118514
 A;Contents: erratum
 A;Accession: A36734
 A;Molecule type: DNA
 A;Residues: 1-1433 <SLO>
 A;Cross-references: GB:M29035; NID:gl43307; PIDN:AAAG2679.1; PID:gl43308
 R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
 J. Bacteriol. 172, 1470-1477, 1990
 A;Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning
 A;Reference number: A35131; MUID:90170864; PMID:2106512
 A;Accession: A35131
 A;Molecule type: DNA
 A;Residues: 1-365,'S',367-682,'EIMP',893,'Q',895-896 <SL2>
 A;Cross-references: GB:M29035
 A;Note: the authors translated the codon GAA for residue 545 as Leu

R;Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
 J. Biol. Chem. 265, 6845-6850, 1990
 A;Title: Cloning, genetic organization, and characterization of a structural gene encoding
 A;Reference number: A35750; MUID:90216713; PMID:2108961
 A;Accession: A35750
 A;Molecule type: DNA
 A;Residues: 1-392,'V',394-828,'NIRTRYSLKFCRSRHKS' <WUA>
 A;Cross-references: GB:J05400; NID:gl42607; PIDN:AAA83362.1; PID:gl42609
 A;Note: this sequence has been corrected
 A;Accession: B35750
 A;Molecule type: DNA
 A;Residues: 876-935,'CG' <WU2>
 A;Cross-references: GB:J05400; NID:gl42607; PIDN:AAA83363.1; PID:gl119197
 A;Note: this sequence has been corrected
 R;Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
 Nucleic Acids Res. 18, 657, 1990
 A;Title: Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus subtilis.
 A;Reference number: S08223; MUID:90174995; PMID:2106671
 A;Accession: S08223
 A;Molecule type: DNA
 A;Residues: 1410-1433 <MAS>
 A;Cross-references: EMBL:X17344; NID:940165; PIDN:CAA35224.1; PID:gb09661
 R;Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
 Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
 A;Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric I
 A;Reference number: JN0335; MUID:93005071; PMID:1368833
 A;Accession: JN0335
 A;Molecule type: protein
 A;Residues: 195-218,'A' <KAT>
 A;Note: source of this material was Bacillus subtilis (natto)
 R;Beall, B.; Lowe, M.; Lutkenhaus, J.
 J. Bacteriol. 170, 4855-4864, 1988
 A;Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli
 A;Reference number: I39846; MUID:89008108; PMID:3139638
 A;Accession: I39849
 A;Status: preliminary; translated from GB/EMBL/DBD
 A;Molecule type: DNA
 A;Residues: 1-211 <RES>
 A;Cross-references: GB:M22630; NID:gl42938; PIDN:AAA22458.1; PID:g551705
 R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berthe
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galles
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
 Kieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrat, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I
 A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: B69596
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-1433 <KUN>
 A;Cross-references: GB:Z99111; GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAB13404.1; P
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: bpr; bpf
 A;Map position: 135 (degrees)
 C;Superfamily: bacillopeptidase F; subtilisin homology
 C;Keywords: extracellular protein; hydrolase; serine proteinase
 F;1-30/Domain: signal sequence #status predicted <SIG>
 F;31-194/Domain: propeptide #status predicted <PRO>
 F;195-1433/Product: bacillopeptidase F #status experimental <MAT>
 F;218-466/Domain: subtilisin homology <SPT>
 F;227,274,452/Active site: Asp, His, Ser #status predicted
 Query Match 5.1%; Score 135.5; DB 1; Length 1433;

Best Local Similarity 20.6%; Pred. No. 1.1;
Matches 112; Conservative 58; Mismatches 150; Indels 223; Gaps 31;

| | | | |
|----|-----|--|-----|
| QY | 90 | ASIPRESF--WFANGTIPAGLYDPEYKV-----PVNADASFSPNFVLDGTAS----- | 136 |
| Db | 437 | SSVPGGTIEDGW--DGTSMAG---PHVSAVALLQKQANASLSDVEMEDILTSTAEPILTDS | 491 |
| QY | 137 | --ADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKT-----YHFTVQRQPGD | 188 |
| Db | 492 | TFPDSNNNGYGHGLVNAFDA-VSAVTDGLGKAEGQVSVEGDDQEPVYQHEKVTEAYEGG | 550 |
| QY | 189 | AASVVVTGEGGNEFAPVQ-----NLQWS-----VSGQTVTILTQAPASD-----KRTY- | 231 |
| Db | 551 | SLPLTLTAEDNVSVTSVKLSYKLDQGEWTEITAKRISGDHLKGTQAEIPDIKGTKLSYK | 610 |
| QY | 232 | -----VLNESPDT-----QTLPGWMTMIDADGDGHNWLSLTINVNT | 267 |
| Db | 611 | WMTHDFGHHVSSDVYDVTVKSIITAGYKQDFETAPGGWV---ASGTNNNW----- | 658 |
| QY | 268 | ATHTGDGAMFSKW-----TASGGAKI---DLSPDNY-----LVTPKVTVPENGK | 309 |
| Db | 659 | -----EWGVPSTGPNPTAASGEKVYGTNLT-GNYANSANMNLVMPPIKAPDSGS | 705 |
| QY | 310 | LSYWYSSQVPWTNEH-----YG-VFLSTTG-----NEAANFTIKLLEETLGSDDKAPMNLVK | 360 |
| Db | 706 | LFLQPKS---MHNLEDDFDYGYVFLPEGEKNWEQAGVYNGK----- | 744 |
| QY | 361 | SEGVKLPAPOERTIDLSAYAGQVYLAFRHFNSTGIER--LYLDDVAVSGSGSNDITY | 418 |
| Db | 745 | -----TSSWTDDEIDLSAYKQNTQVMENLOSDESIAKEGWIIDVVLSDKSAGK---- | 794 |
| QY | 419 | TVYRDNVVI-----AQNLAAATTFQENVA----- | 442 |
| Db | 795 | TVKKNKLGVEKPSGKQKKKPVNPKKAKPSANTAVKHQKAIQPVLPKAAQVSVVETGKS | 854 |
| QY | 443 | -----PGQYNYCVVKYTAGVSPKVKCDVTVEGNEFAHVQNLTGSAVG-----OKVTL | 491 |
| Db | 855 | TVSDQSTGQYT---LKHKAG-----DYT-----LMAEAYGYQSKTKVSL | 891 |
| QY | 492 | KWD 494 | |
| Db | 892 | KTD 894 | |

Search completed: May 18, 2004, 11:47:45
Job time : 15.9765 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 9.08355 Seconds
(without alignments)
2848.981 Million cell updates/sec

Title: US-08-570-311-2

Perfect score: 2641

Sequence: 1 MRKNSLFLSLVLLSLCWG.....QNLTSAGVQKVTWKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 2641 | 100.0 | 2628 | 1 | HGA2_PORGI |
| 2 | 2583 | 97.8 | 2164 | 1 | HGA2_PORGI |
| 3 | 494.5 | 18.7 | 991 | 1 | CPGI_PORGI |
| 4 | 472 | 17.9 | 989 | 1 | PRTH_PORGI |
| 5 | 156 | 5.9 | 1157 | 1 | XYNA_THESA |
| 6 | 137.5 | 5.2 | 635 | 1 | XYND_PABPO |
| 7 | 135.5 | 5.1 | 1433 | 1 | SUBF_BACSU |
| 8 | 135 | 5.1 | 1655 | 1 | OMPB_RICCN |
| 9 | 131.5 | 5.0 | 725 | 1 | YA33_SULSO |
| 10 | 131 | 5.0 | 1656 | 1 | OMPB_RICJA |
| 11 | 129 | 4.9 | 794 | 1 | OXAA_CHLCV |
| 12 | 127 | 4.8 | 2710 | 1 | TOXA_CLODI |
| 13 | 124 | 4.7 | 1616 | 1 | SLAP_BACCI |
| 14 | 121.5 | 4.6 | 972 | 1 | CTAL_BACCI |
| 15 | 120.5 | 4.6 | 587 | 1 | PICP_PSSSR |
| 16 | 119 | 4.5 | 699 | 1 | CHII_BACCI |
| 17 | 118.5 | 4.5 | 1300 | 1 | 120K_RICRI |
| 18 | 118.5 | 4.5 | 1654 | 1 | OMPB_RICRI |
| 19 | 115.5 | 4.4 | 336 | 1 | G3P1_SCHPO |
| 20 | 114.5 | 4.3 | 335 | 1 | G3P1_GRAVE |
| 21 | 114.5 | 4.3 | 488 | 1 | PHB_ALCPA |
| 22 | 114.5 | 4.3 | 681 | 1 | TKT2_YEAST |
| 23 | 114.5 | 4.3 | 857 | 1 | GELA_DICDI |
| 24 | 114.5 | 4.3 | 2660 | 1 | YEBJ_ECO57 |
| 25 | 113.5 | 4.3 | 1004 | 1 | SLPO_BACBR |
| 26 | 113.5 | 4.3 | 1151 | 1 | ITAL_HUMAN |
| 27 | 111 | 4.2 | 2201 | 1 | TENA_HUMAN |
| 28 | 111 | 4.2 | 2358 | 1 | YEEJ_ECOLI |
| 29 | 110.5 | 4.2 | 408 | 1 | G3PT_HUMAN |
| 30 | 110.5 | 4.2 | 2265 | 1 | FINC_BOVIN |
| 31 | 110 | 4.2 | 642 | 1 | FLID_CAME |
| 32 | 110 | 4.2 | 827 | 1 | CSG_HALVO |
| 33 | 110 | 4.2 | 1409 | 1 | HAPI_HAEIN |

34 109.5 4.1 932 1 CDG3_HUMAN
35 109.5 4.1 2021 1 OMPA_RICCN
36 109 4.1 591 1 FLGE_CAUCR
37 109 4.1 917 1 SLAP_THETH
38 109 4.1 1120 1 STFR_ECOLI
39 108.5 4.1 337 1 G3P_PODAN
40 108 4.1 611 1 WDRI_CABEL
41 108 4.1 828 1 MRKC_KLEPN
42 108 4.1 2481 1 FINC_XENLA
43 107.5 4.1 1475 1 APU_THETI
44 107.5 4.1 2477 1 FINC_RAT
45 107 4.1 666 1 NEPI_THEVU

ALIGNMENTS

RESULT 1
HGA2_PORGI
ID HGA2_PORGI STANDARD; PRT; 2628 AA.
AC Q51845;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin A precursor.
GN HAGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97047672; PubMed=8926061;
RA Han N., Whitlock J., Frogulske-Fox A.;
RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
contains four large, contiguous, direct repeats.";
RL Infect. Immun. 64:4000-4007(1996).
CC -!- FUNCTION: Agglutinates erythrocytes.
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; U41807; AAB17128.1; --
PIR; T28651; T28651.
KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT SIGNAL 1 24 POTENTIAL:
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 100.0%; Score 2641; DB 1; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.1e-162;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRKNSLFLSLVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVKDLRDP 60
Db 1 MRKNSLFLSLVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVKDLRDP 60
Qy 61 ARTILRAHDWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120
Db 61 ARTILRAHDWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120

QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVYVNESFDTQ 240
 DB 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVYVNESFDTQ 240
 QY 241 TLPNGWTMIDADGHNWLSLTIINVTATHTGDMFSGKSWTASGAKIDLSPDNYLVT 300
 DB 241 TLPNGWTMIDADGHNWLSLTIINVTATHTGDMFSGKSWTASGAKIDLSPDNYLVT 300
 QY 301 KVTVPENGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLBETLGSKDPAPMNLVK 360
 DB 301 KVTVPENGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLBETLGSKDPAPMNLVK 360
 QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYTV 420
 DB 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYTV 420
 QY 421 YRDNVVIAQNLAAATTFNOENVAPQVNYCVBKVTAGVSPKCKDVTVEGSEFAHVQNL 480
 DB 421 YRDNVVIAQNLAAATTFNOENVAPQVNYCVBKVTAGVSPKCKDVTVEGSEFAHVQNL 480
 QY 481 TGSVAVGQKVTWKWDAPN 497
 DB 481 TGSVAVGQKVTWKWDAPN 497

RESULT 2

ID HGAI PORGI STANDARD; PRT; 2164 AA.
 AC P5915;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hemagglutinin A precursor.
 GN HAGA OR PG1837.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=22829867; PubMed=12949112;
 RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
 RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
 RA Dewhirst F.E., Fraser C.M.;
 RT "Complete genome sequence of the oral pathogenic bacterium
 Porphyromonas gingivalis strain W83.";
 RL J. Bacteriol. 185:5591-5601(2003).
 CC -!- FUNCTION: Agglutinates erythrocytes (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family C25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AE017178; AAQ66831.1; ALT_INIT.
 DR TIGR; F61837; -.
 KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2164 HEMAGGLUTININ A.
 FT DOMAIN 26 539 PEPTIDASE C25-LIKE 1.
 FT DOMAIN 540 991 PEPTIDASE C25-LIKE 2.

FT DOMAIN 992 1443 PEPTIDASE C25-LIKE 3.
 SQ SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;

Query Match 97.8%; Score 2583; DB 1; Length 2164;
 Best Local Similarity 98.0%; Pred. No. 4.9e-159;
 Matches 407; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 MRKLSLPSLAVLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVKDLRDPAPGM 60
 DB 1 MRKLSLPSLAVLSLLCWGQTAAAGGPKTAPSVTHQAVQKGIKTSKVKDLRDPAPGM 60
 QY 61 ARITLEAHVDWEDGTGYQMLADHNOYASIPRESFWFANGTIPAGLYDPEYKVPVNA 120
 DB 61 ARITLEAHVDWEDGTGYQMLADHNOYASIPRESFWFANGTIPAGLYDPEYKVPVNA 120
 QY 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180
 DB 121 DASFSPTNFVLDGTASADIPAGTYDYVIINPNPGIYIIVGEGVSKGNDYVVEAGKTYHFT 180
 QY 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVYVNESFDTQ 240
 DB 181 VORQPGDAASVVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVYVNESFDTQ 240
 QY 241 TLPNGWTMIDADGHNWLSLTIINVTATHTGDMFSGKSWTASGAKIDLSPDNYLVT 300
 DB 241 TLPNGWTMIDADGHNWLSLTIINVTATHTGDMFSGKSWTASGAKIDLSPDNYLVT 300
 QY 301 KVTVPENGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLBETLGSKDPAPMNLVK 360
 DB 301 KVTVPENGKLSYVSSQVPTWNEHYGVFLSTTGNEAANFTIKLBETLGSKDPAPMNLVK 360
 QY 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYTV 420
 DB 361 SEGKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDVAVSGEGSSNDYTYTV 420
 QY 421 YRDNVVIAQNLAAATTFNOENVAPQVNYCVBKVTAGVSPKCKDVTVEGSEFAHVQNL 480
 DB 421 YRDNVVIAQNLAAATTFNOENVAPQVNYCVBKVTAGVSPKCKDVTVEGSEFAHVQNL 480
 QY 481 TGSVAVGQKVTWKWDAPN 497
 DB 481 TGSVAVGQKVTWKWDAPN 497

RESULT 3

ID CP61 PORGI STANDARD; PRT; 991 AA.
 AC P28784; Q45168;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
 DE (RGP-1).
 GN RGP1.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
 RC STRAIN=381;
 RX MEDLINE=95168884; PubMed=7864651;
 RA Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
 RA Ikehara Y.;
 RT "Structural characterization of argingipain, a novel
 arginine-specific cysteine proteinase as a major periodontal
 pathogenic factor from Porphyromonas gingivalis.";
 RL Arch. Biochem. Biophys. 316:917-925(1995).
 RN [2]
 RP SEQUENCE OF 228-270.
 RC STRAIN=HG66;
 RX MEDLINE=92406812; PubMed=1527017;
 RA Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;

RT "Purification and characterization of a 50-kDa cysteine proteinase
 RL (gingipain) from Porphyromonas gingivalis";
 CC J. Biol. Chem. 267:18896-18901(1992).
 CC -!- FUNCTION: Thiol protease which is believed to participate in
 CC intracellular degradation and turnover of proteins. Its
 CC proteolytic activity is a major factor in both periodontal tissue
 CC destruction and in bacterial host defense mechanisms. Activates
 CC complement C3 and C5.
 CC -!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and
 CC immunoglobulins, with a preference for Arg in P1, and hydrophobic
 CC residues in P2 and P3.
 CC -!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+)
 CC and/or Mg(2+) for stabilization. It is stimulated by glycine-
 CC containing dipeptides. It is resistant to inhibition by proteinase
 CC inhibitors in human plasma.
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D26470; BAA05484.1; -.
 CC PIR; I40229; I40229.
 CC HSP; P95493; 1CVR.
 CC MEROPS; C25.001; -.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR001769; Peptidase_C25.
 CC InterPro; IPR005536; Peptidase_C25_C.
 CC Pfam; PF01364; Peptidase_C25; I.
 CC Pfam; PF03785; Peptidase_C25; C; 1.
 CC Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
 CC SIGNAL 1 24
 CC PROPEP 25 227
 CC CHAIN 228 991 GINGIPAIN R1.
 CC ACT_SITE 438 438 PROTON DONOR (BY SIMILARITY).
 CC ACT_SITE 471 471 NUCLEOPHILE (BY SIMILARITY).
 CC CONFLICT 264 265 RT -> TK (IN REF. 2).
 CC SEQUENCE 991 AA; 108782 MW; 03EE3F43CEBE2544 CRC64;
 CC
 CC Query Match 18.7%; Score 494.5; DB 1; Length 991;
 CC Best Local Similarity 25.5%; Pred. No. 2e-24;
 CC Matches 124; Conservative 41; Mismatches 90; Indels 231; Gaps 7;
 CC
 CC QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVKOLRDIPAGMARILIEAH 68
 CC Db 671 TATTGQKVTWKWDAFSTKTNTATNTARSVDGIRLVLLSVSDAPELRSGQAEIVLEAH 730
 CC
 CC QY 69 DVWEDGTGYQLMDADHNOYGASIPER--SFWFANGTIPAGLYDPREYKVPVNADASFSP 126
 CC Db 731 DVWNGSGYQLLDADHDQGVIPSTHILW--PNCSPANLPAFFYTYFENADPSCSP 789
 CC
 CC QY 127 TNFVLDGTASADIPAGTYDYVVIINPNFI--IYIVGEGVSGNDYVVEAGKTYHFTVROQG 185
 CC Db 790 TNMIMDGTASVNIAGTYDFALAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKMG 849
 CC
 CC QY 186 PGDAASVVVTGEGNEFAPYQNLQWSVGGTIVLTWQAPASDKRTYVINESFDQTLPNG 245
 CC Db 850 SGGDTELTIS----- 859
 CC
 CC QY 246 WTMDADGDGHNWLSTINVTYNTATHTGDGAMFSKSWTASGAKIDLSPDNLYVTPKVTVP 305
 CC Db 860 ----- 859
 CC
 CC QY 306 ENGLSYWSSQVPWNEHYGVFLSTTGNENAFNIKLBETLGLSKDPAPMNLVKGSGVK 365
 CC Db 860 ----- 859
 CC
 CC QY 366 LPAPYQERTIDLSAYAGQVYLAFPHNENGTIFRLYLDVAVSGEGSSNDYTYVYRDNV 425
 CC Db 366 LPAPYQERTIDLSAYAGQVYLAFPHNENGTIFRLYLDVAVSGEGSSNDYTYVYRDNV 425

Db 860 -----EGGSDYTYTVYRDGT 875
 QY 426 VIAQLAATFTNQENVAPGQYNYCYEYKTAGVSPKCKCVTVEGSNEFAHVQLTGSAY 485
 Db 876 KIKELGTETTYRDAGMSAQSHCYEVKRYAAGVSPKCVDPYIPDGVADVTAKPYTLTW 935
 QY 486 GQKVTLL 491
 Db 936 GKTITV 941
 RESULT 4
 PRTH PORGI
 ID PRTH PORGI STANDARD; PRT; 989 AA.
 AC P46071;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Protease prth (EC 3.4.22.-).
 DB PRTH
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 CC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 CC Porphyromonadaceae; Porphyromonas.
 CC NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=95012612; PubMed=7927685;
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RT "Cloning and characterization of a new protease gene (prth) from
 RL Porphyromonas gingivalis";
 RN Infect. Immun. 62:4279-4286(1994).
 RN [2]
 RP ERATUM.
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RL Infect. Immun. 62:5707-5707(1994).
 CC -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
 CC P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
 CC IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
 CC AND IS A VIRULENCE FACTOR.
 CC -!- SUBCELLULAR LOCATION: In membrane vesicles.
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L27483; AAA51298.1; -.
 CC HSP; P23882; IFMT.
 CC MEROPS; C25.001; -.
 CC InterPro; IPR002376; formyl transf.
 CC Pfam; PF00551; formyl transf. 1.
 CC Hydrolase; Thiol protease; Repeat; Virulence.
 CC REPEAT 270 323
 CC REPEAT 528 581
 CC SEQUENCE 989 AA; 110238 MW; FA85FE8A3AC8944C CRC64;
 CC
 CC Query Match 17.9%; Score 472; DB 1; Length 989;
 CC Best Local Similarity 38.6%; Pred. No. 5.7e-23;
 CC Matches 119; Conservative 45; Mismatches 124; Indels 20; Gaps 10;
 CC
 CC QY 199 GNEFAPVQLWSVSGTIVLTWQAPASDKRTYVINESFDQT---LPNGWTMIDADGDG 255
 CC Db 58 GTEICCFPSLPKAPICSTSLMLRSKTNKRAADFTETESSHGCAPAEWTIDADGDG 117
 CC
 CC QY 256 HNW--LSTINVTYNTATHTGDGAMFSKSWTASGAKIDLSPDNLYVTPKVTVPNGKLSY 313
 CC Db 118 QGWLCLSSQLDMLTAHGGTNNVSPSWNG-----MALNPNDYLSKDVGTGATKYYVA 172

Science 293:2093-2098 (2001).
 [2]
 SEQUENCE OF 33-1649 FROM N.A.
 STRAIN=Indian tick typhus and Malish 7;
 MEDLINE=20393643; PubMed=10939643;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 gene coding the outer-membrane protein rOmpB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 RN [3]
 RP SEQUENCE OF 353-1655 FROM N.A.
 RC STRAIN=Malish 7;
 RA Stenos J., Walker D.;
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia
 australis, the most divergent rickettsia of the spotted fever group.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases
 CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
 CC similarity).
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE008659; AAL03623.1; -
 CC EMBL; AF123721; AAF34124.1; -
 CC EMBL; AF123726; AAF34129.1; -
 CC EMBL; AF149110; AAD39533.1; -
 CC PIR; E97835; E97835.
 CC InterPro; IPR006315; Autotransport.
 CC InterPro; IPR005546; Autotransporter.
 CC Pfam; PF03797; Autotransporter; 1.
 CC TIGRams; TIGR01414; autotrans parl; 2.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1335 1655 32 kDa BETA PEPTIDE.
 FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 353 354 KD -> GH (IN REF. 3).
 FT CONFLICT 776 776 F -> S (IN REF. 3).
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).
 SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;
 Query Match 5.1%; Score 135; DB 1; Length 1655;
 Best Local Similarity 19.3%; Pred. No. 0.7;
 Matches 102; Conservative 53; Mismatches 189; Indels 184; Gaps 20;
 QY 100 ANGTFAGLYDPPEYKYPVNVADASFSTNFV-----LDGTASADIAGTYDVY 147
 Db 439 ANGLTAS-----ASADANVANVTNNITAEASGAGVQVLSGTHAAELRLG----- 482
 QY 148 IINPNPGIIVIGEVG---SKGNDYVVEAGKTHFTVORQPGDAASVVVTGEGNE--F 202
 Db 483 ---NAGSVFKLADGTGVNGKVNQVLTALVGGAALAGTITLDG-----SATITGIGNAGGA 533

203 APVO-----NLQWSVSGQTVTLTWQ-----APASD 227
 Db 534 AALQGITLANDATKTLTGGANIIGANGGTINFOAGGTIKLTSTONNIVVDFDLATAD 593
 QY 228 KRTYVLNESFDQTLPNGWTM-----IDADDGHNWLSLNNVNTATHTATGDMFSSKSWT 282
 Db 594 QTGVW-----DASSLTNAQTLTINGKGTGVGANNKTLGQFNIGSSKTVLSGDVAINELV 648
 QY 283 ASGAKIDLSPDNYLVTPKVTPENGKLSVWSSQVPTWNEHYGVFLSTTGCNEAANFTIK 342
 Db 649 IGNNGAVQFAHTYLIITRTTNAAGQGIIF-----NFWVNNNTTL-----ATGTLGSAATNP 700
 QY 343 LLEBETLSDKPPAPMLVKS--EGVKLPAPQERTIDLSAVAGQOVYLAFRHFNSTGLFRL 400
 Db 701 LAEINFSGKAANVDVTVNVKGVNL---YATNITTTDANVGSFIFA---GGTNI--- 750
 QY 401 YLDDVAVSGE--GSSNDYTYTVYRDNVVIAQNLAATTFNQENVAPGQVNYCDEVKYTA-- 456
 Db 751 -----VSGTVGGQGNKENTVALDNGTIVKELGNATFNNGTTIAANSTLIQIGGNYTADF 804
 QY 457 -----GVSPKVKDVTVEGS-----NEFAHVQNLTG----- 482
 Db 805 VASADGTGIVEFVNTGPIITVLNKQAAPVNALKQITVSGFNVVINEIGNAGNYHGAATD 864
 QY 483 -----SAVGOKVTLKWDAPN 497
 Db 865 TIAFENSSLGAVFLPRGIFPDNAGNRIPITIKSTVGKNTATGDFVPS 912
 RESULT 9
 YA33 SULSO
 ID YA33 SULSO STANDARD; PRT; 725 AA.
 AC Q97Z57;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein SSO1033 precursor.
 GN SSO1033.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 EX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -!- SIMILARITY: Contains 3 fibronectin type III domains.
 CC -!- SIMILARITY: Contains 6 Kelch repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE006722; AAK41296.1; -
 CC PIR; A90255; A90255.
 CC InterPro; IPR008957; FN III-like.
 CC InterPro; IPR003961; FN_III.
 CC InterPro; IPR006632; Kelch_rep.
 CC Pfam; PF00041; fn3; 2.
 CC Pfam; PF01344; Kelch; 5.

DR SMART; SM00060; FN3; 3.
 KW Hypothetical protein; Kelch repeat; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 28
 FT CHAIN 29 725
 FT REPEAT 53 100
 FT REPEAT 101 145
 FT REPEAT 146 199
 FT REPEAT 201 248
 FT REPEAT 250 297
 FT REPEAT 299 342
 FT DOMAIN 323 398
 FT DOMAIN 412 490
 FT DOMAIN 585 652
 SQ SEQUENCE 725 AA; 78465 MW; FE8F1220D013BAB6 CRC64;

Query Match 5.0%; Score 131.5; DB 1; Length 725;
 Best Local Similarity 20.4%; Pred. No. 0.4; Indels 241; Gaps 27;
 Matches 112; Conservative 61; Mismatches 134; Indels 241; Gaps 27;

QY 101 NGTIPAGLYDPPEYKVPVNADASFPTN-----FVLDTASADI----- 139
 Db 154 NNTSPAGLYFP-----PSNAIRLFYPNNDWRILGYMPVPTYGGVFNGLTSLIIVSGVI 208
 QY 140 --PAGTYDVIINPN-----PGIIVYEGVSKGNDYVVEAG 174
 Db 209 GYSAYTNDILYSPQNNWNTEILNGVLPYWIHDSALAYRGVLFVIG-----GYIYTAG 261
 QY 175 K-----TYHFTVQROG-----PGDAASVY-----VTEGEGNEFAPVQNLQ----- 209
 Db 262 SGVANNAILAYNGNLQRYGLYLPVYYSAGYVQVGNMLYAGIGSSISDVSAQLITFN 321
 QY 210 -----NSVSGQTVTLTWQAPASDKRTYVLNSESFTQTLPRGWTMIDADGQHNWL 259
 Db 322 FPPLPKITSYAGNESVTLGW-----NEVRLSSGYEIIYNNWGFN-- 363
 QY 260 STINYNTATHI-----GDCAMF----- 277
 Db 364 SSINVGNTSYVTGKDGITYFYFVLAYNSIGYSPSSIIALTASVNPPLQVSVRYK 423
 QY 278 ----SKSW-----TAGS-----GAKIDLSPDNVLY-----TPKVT----- 303
 Db 424 NDNVTLNMLPFTFSGYLLGYVIVKNSNMVSHFNSTSLTTSNTPNVTYVNFYIA 483
 QY 304 VPENKLSYVWSSQVPWNEHYGVFLSTGNE-AAAFITKLEETLGSDKPKAPMLVKSE 362
 Db 484 VNKLGNSSPLVLTVPVITKASVFATITKLGNLILVNWITSF-----PANIT-- 529
 QY 363 GVKLPAQERTIDLSAYAGQVYLAFR-----HFNSTGIFR-----LYL-- 402
 Db 530 -LELYNPNGNLISQIAATKGNSSYL-FRPVQGNVTLVIIASNSAGVSKYVQVYVYLP 587
 QY 403 -----DDVAVSGGSSNDYTYTVYVRDNNVIAQNLAAATTFNQENVAPQYNYCEV 452
 Db 588 SPQVSLIGFNGNLIXSWNEANVITLYVYVNNSLVYEGFSPNSIVT--NISNG--TYLVKV 643
 QY 453 KYTAGVSP 460
 Db 644 ---IGVNP 648

RESULT 10
 OMPB_RICJA STANDARD; PRT; 1656 AA.
 AC 006653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOmpB)
 DE (rOmp B) [contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB.
 OS Rickettsia japonica.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia japonica."
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB003681; BAA20138.1; -;
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR005546; Autotransporter.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRfams; TIGR01414; autotrans_bar1; 2.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 5.0%; Score 131; DB 1; Length 1656;
 Best Local Similarity 21.3%; Pred. No. 1.3;
 Matches 99; Conservative 45; Mismatches 198; Indels 122; Gaps 18;

QY 100 ANGTPAGLYDPPEYKVPVNADASFPTNFV-----LDGTASADIPAGT--- 143
 Db 439 ANGTLAS-----ASADANVAVTNNTITAEASGVGVQLSGTHRELRLGNAGSV 487
 QY 144 ---YDYVIINPNPGIIVYEGVSKGNDYVVEAGKTYHFTVQROGP-----DAA 190
 Db 488 FKLADGTGVINGKVNQTVLVG-GVLAAGAITLDGSATITGDIGNGGGAALQSITLANDAT 546
 QY 191 SVVVTGEGNEFAPVQNLQWSVSGQVTLTWQ-----APASDKRTYVLNSESFTQ 240
 Db 547 KTLTLGGANITISANGTINFGANGTITKLTSTQNNIVVDCDLATATDQTGVV-----DAS 601
 QY 241 TLPNGWTM-----IDADGDGHNMLSTINVTATHTGCGAMFSKSWTASGGAKIDLSPDN 295
 Db 602 SLTNAQTILTSGTIGIIGANNITLIGQFNIGSSKTLINGNVAINELVINGNSVQFAHNT 661
 QY 296 YLVTVPKVTVPENKLSYVWSSQVPWNEHYGVFLSTTGNBAANPTIKLLETGLSDKPKAP 355
 Db 662 YLITRTTNAAGQGGKIIF-----NPVVNNNTTLAAGTNLGSAAAN---PLAEIFGS-KGAR 712
 QY 356 MNLVKS--EGVKLPAPQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDLDDVAVSGE--G 411
 Db 713 ADTVLVNVEGVNL---YATNITTTDANVGSFV-----FNAG-----KNIVSGTVGG 756
 QY 412 SSNDYTYTVVRDNNVIAQNLAAATTFNQENVAPQYNYCEVVKYTA----- 456
 Db 757 QQGNKFNVALDNGTTVKFLGNATFNQNTIAANSTLQISGNTYADFIASADGTGIVEFV 816
 QY 457 -----GVSPKVKCDVTVEGS-----NEFAHVQNLTG 483

Db 817 NTGPINVTLNKQAVFVNALKOITVSGPGNVVVVNEIGNAGNYHGA 860

RESULT 11

OXAA_CHLCV STANDARD; PRT; 794 AA.
AC PS9809;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Inner membrane protein oxaA.
GN OXAA OR CCA00472.
OS Chlamydomophila caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12692364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzaple E., Khouri H., Federova N.B., Carty H.A.,
RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoll P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
examining the role of niche-specific genes in the evolution of the
Chlamydiaceae.";
RL Nucleic Acids Res. 31:2134-2147(2003).
CC -|- FUNCTION: Required for the insertion of integral membrane proteins
into the membrane. Probably plays an essential role in the
integration of proteins of the respiratory chain complexes.
CC Involved in integration of membrane proteins that insert
independently and independently of the Sec translocase complex (By
similarity).
CC -|- SUBUNIT: Specifically interacts with transmembrane segments of
nascent integral membrane proteins during membrane integration (By
similarity).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(By similarity).
CC -|- SIMILARITY: Belongs to the OXAL/oxaA family. Subfamily 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016995; AAP05217.1; -.
CC TIGR; CCA00472; -.
CC HAMAP; MF 01810; -; 1.
DR InterPro; IPR001708; 60kDa innermem.
DR Pfam; PF02096; 60KD IMP; 1-
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 5 24 POTENTIAL.
FT TRANSMEM 571 593 POTENTIAL.
FT TRANSMEM 639 661 POTENTIAL.
FT TRANSMEM 691 713 POTENTIAL.
FT TRANSMEM 740 762 POTENTIAL.
SQ SEQUENCE 794 AA; 88951 MW; 9EDCF05E73709133 CRC64;

Query Match

Best Local Similarity 4.9%; Score 129; DB 1; Length 794;
Matches 86; Conservative 63; Mismatches 154; Indels 106; Gaps 21;
QY 80 LWADHNOYCASPEESFW-FANGTTPAGLYDPEYK-----VPVNADAFSPTNFVLDG 133
Db 167 VYNKDSAVYGTSL---VFWRSGNEYLPLGTYNGKEERLESIDLPLTKAAVFSDSKSNAG 223
QY 134 TASAD--IPAGTYDYVNIINPGLIYVGEV-----SKGNDYVW-FAGTYHTVVRQG 195
Db 224 ANSAQFVLSNENYQIVSOGESI-----EGINLPFSSSEDNKSIVNEIG--FDRELKAQV 277

QY 186 PGDAASVVVTCEGGNEFAPYONLWVSQGVTVTLTWQAPASD--KRTYVLNESFDQTLP 243
Db 278 PSEAFPLPSVGGANN-QPVS-----TVGGYPLRRGILSDAKKRT---PSSYHALNIV 329
QY 244 NGWTMIDADGDGHNLSTINVTATHTGCGAM-----FSSKS 280
Db 330 SGRELTSVASGYR-VSTFSTMLESDNGSIKKYKLPQOQYAFVEVGVNRSDDL 388
QY 281 WTASGGAKIDLSPDNYLVTP-----KVTVE-----NGKLSYVWSSQ 317
Db 389 WITSGIPEVEIMSNF--TPAIKYHVIVKKNKGQLDKVLKPKADPLALRSVGVPQWILN- 445
QY 318 VPWTHYGVFLSTGTGNEAANFTIKLEETIGSDKPAFNMVLKSEG-----VK 365
Db 446 ---SNGYFGIILSLPTDIPAGYAAAYVP---GSSVPTRLSLSPKNOAYPASKYPGYETL 499
QY 366 LPAPYQERTIDLSAYAGQOQVYLAFR-----HFNSTGIFRLYLDVAVSG 409
Db 500 LPLPQKSGTHRFVYAGPLADPTLRVLDKAYTNSKGESPOYLDCITFRG 548
RESULT 12
TOXA_CLODI STANDARD; PRT; 2710 AA.
AC PL6154;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90221894; PubMed=2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A.";
RL Nucleic Acids Res. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=30129305; PubMed=2105276;
RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyster D.M.,
RA Wilkins T.W., Johnson J.L.;
RT "Molecular characterization of the Clostridium difficile toxin A
gene.";
RL Infect. Immun. 58:480-488(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: Only after the enteral delivery of the enterotoxin A may
the characteristic disease called pseudomembranous colitis be
induced.
CC -|- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
DIFFERENT OLIGOPEPTIDES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X51797; CAA36094.1; -.
CC EMBL; M30307; AAA23283.1; -.
DR

KW Hydrolase; Protease; Serine protease; Zymogen; Periplasmic; Signal;
KW 3D-structure; Calcium-binding.

FT SIGNAL 1 ? POTENTIAL.
FT PROPEP 2 215 REMOVED IN MATURE FORM.
FT CHAIN 216 585 PSEUDOMONALISIN.
FT PROPEP 586 587 REMOVED IN MATURE FORM.
FT ACT SITE 295 299 CHARGE RELAY SYSTEM.
FT ACT SITE 299 239 CHARGE RELAY SYSTEM.
FT ACT SITE 502 502 CHARGE RELAY SYSTEM.
FT DISULFID 352 391
FT METAL 543 543 CALCIUM.
FT METAL 544 544 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 559 559 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 561 561 CALCIUM (VIA CARBONYL OXYGEN).
FT METAL 563 563 CALCIUM.
SQ SEQUENCE 587 AA; 61072 MW; E193DBB2C225829A CRC64;

Query Match 4.6%; Score 120.5; DB 1; Length 587;
Best Local Similarity 23.5%; Pred. No. 1.6;
Matches 78; Conservative 33; Mismatches 108; Indels 113; Gaps 16;

QY 103 TIPAGL---YDPFEYKVPWNADAFSPPTNFVLDTGTASADIPAGTYDYVYIINENPGIYIV 159
Db 214 TLAAGTAKHNPTEF--PTIYDASSAPT-----AANTVGIITIG 251
QY 160 G-----EGVSKGNDYVVEAGKT---YHFTVQRQGGDAASVVVTGEGGNEF 202
Db 252 GVSQTLQDLQOFTSANGLASVNTQITGSSNGDYSDQGGGEWDLDSQIVGSAG--- 308
QY 203 APVONLOW-----SVSGQT-VTLTWQAPASDKRTYVLNESFDQTLPNGWTMIDADGDGH 256
Db 309 GAVQQLLFYMDQDSAGSNTGLTQAFNAQVSDNVAKVINVSL-----GWCEADANADG- 360
QY 257 NWLSTINVTATHTG-----DGAMFSKSWTASGGAKIDLSPDN 295
Db 361 TLQEDRIFATAAGQGTFSVSSGDEGVYECNRCYPDGTYSVSWPAS-----SPNV 413
QY 296 YLVTPKVTVPENGLSYVWSSQVPWNHGYVFLSTTCNEAANFTIKLLEETLG-----SD 351
Db 414 IAV-----GGTTLTYTSAGAYSNE-----TVWNEGLDSNGKLWATGGGVSVYES 457
QY 352 KPAPMNLVKSQGVKLPAFYQERTIDLSAYAQ 383
Db 458 KPSWQSVVSG-----TGRRLLPDISPDAQ 483

Search completed: May 18, 2004, 11:43:30
Job time : 10.2836 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:36:20 ; Search time 36.3342 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-2

Perfect score: 2641

Sequence: 1 MRKLSLPSLAVLLSLCWG.....QNLGSAVGQVKTLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | ID | Description |
|------------|-------|---------|--------------|--------|--------------------|
| 1 | 951.5 | 36.0 | 1706 | Q51839 | Q51839 porphyromon |
| 2 | 950.5 | 36.0 | 1706 | Q51838 | Q51838 porphyromon |
| 3 | 944.5 | 35.8 | 1704 | Q51816 | Q51816 porphyromon |
| 4 | 943.5 | 35.7 | 1697 | Q9R9B7 | Q9R9B7 porphyromon |
| 5 | 824 | 31.2 | 1723 | Q72194 | Q72194 porphyromon |
| 6 | 818 | 31.0 | 1723 | Q72197 | Q72197 porphyromon |
| 7 | 816 | 30.9 | 1358 | Q96967 | Q96967 porphyromon |
| 8 | 812.5 | 30.8 | 1732 | Q51817 | Q51817 porphyromon |
| 9 | 807.5 | 30.6 | 1732 | Q52050 | Q52050 porphyromon |
| 10 | 806.5 | 30.5 | 1732 | Q07442 | Q07442 porphyromon |
| 11 | 805.5 | 30.5 | 1223 | Q9ZNB5 | Q9ZNB5 porphyromon |
| 12 | 561.5 | 21.3 | 1097 | Q72196 | Q72196 porphyromon |
| 13 | 379 | 14.4 | 312 | Q9KIB3 | Q9KIB3 porphyromon |
| 14 | 337 | 12.8 | 925 | Q9F4J0 | Q9F4J0 porphyromon |
| 15 | 213.5 | 8.1 | 293 | Q9XBU9 | Q9XBU9 porphyromon |
| 16 | 156 | 5.9 | 1454 | Q8ENZ2 | Q8ENZ2 oceanobacil |

| | | | | | | |
|----|-------|-----|------|----|--------|---------------------|
| 17 | 154 | 5.8 | 1155 | 17 | Q8TPZ8 | Q8tpz8 methanosarc |
| 18 | 153 | 5.8 | 5291 | 16 | Q8X2T1 | Q8x2t1 escherichia |
| 19 | 152 | 5.8 | 1234 | 2 | Q60046 | Q60046 thermoanaer |
| 20 | 150 | 5.7 | 1483 | 16 | Q977Y4 | Q977y4 clostridium |
| 21 | 147 | 5.6 | 1983 | 16 | Q9CJ55 | Q9cj55 lactococcus |
| 22 | 146 | 5.5 | 2951 | 17 | Q8THC9 | Q8thc9 methanosarc |
| 23 | 144.5 | 5.5 | 781 | 16 | Q9S2G1 | Q9s2g1 streptomyce |
| 24 | 139 | 5.3 | 5020 | 16 | Q8E9W3 | Q8e9w3 shewanella |
| 25 | 138.5 | 5.2 | 2468 | 16 | Q9I2M3 | Q9i2m3 pseudomonas |
| 26 | 137.5 | 5.2 | 8173 | 16 | Q7UDU8 | Q7udu8 rhodopirell |
| 27 | 137 | 5.2 | 2013 | 16 | Q92EK2 | Q92ek2 listeria in |
| 28 | 135 | 5.1 | 1615 | 2 | Q9F0P9 | Q9f0p9 rickettsia |
| 29 | 135 | 5.1 | 1654 | 2 | Q93QW9 | Q93qw9 rickettsia |
| 30 | 135 | 5.1 | 3064 | 16 | Q82XT8 | Q82xt8 nitrosomona |
| 31 | 133 | 5.0 | 897 | 16 | Q8G7K7 | Q8g7k7 bifidobacte |
| 32 | 133 | 5.0 | 1734 | 17 | Q8TI67 | Q8ti67 methanosarc |
| 33 | 132 | 5.0 | 698 | 16 | Q82F44 | Q82f44 streptomyce |
| 34 | 131.5 | 5.0 | 705 | 16 | Q93J50 | Q93j50 streptomyce |
| 35 | 131.5 | 5.0 | 768 | 16 | Q88G76 | Q88g76 pseudomonas |
| 36 | 131.5 | 5.0 | 1379 | 17 | Q9HL55 | Q9hl55 thermoplasma |
| 37 | 131.5 | 5.0 | 1618 | 2 | Q9KKB4 | Q9kkb4 rickettsia |
| 38 | 131 | 5.0 | 1357 | 17 | Q8THC8 | Q8thc8 methanosarc |
| 39 | 131 | 5.0 | 1618 | 2 | Q9KKB1 | Q9kkb1 rickettsia |
| 40 | 131 | 5.0 | 2009 | 16 | Q7UY44 | Q7uy44 rhodopirell |
| 41 | 131 | 5.0 | 2768 | 16 | Q8E9G6 | Q8e9g6 shewanella |
| 42 | 130.5 | 4.9 | 524 | 16 | Q7UWU6 | Q7uwu6 rhodopirell |
| 43 | 130.5 | 4.9 | 1616 | 2 | Q9KKB0 | Q9kkb0 rickettsia |
| 44 | 130.5 | 4.9 | 1617 | 2 | Q9KKB9 | Q9kkb9 rickettsia |
| 45 | 130.5 | 4.9 | 2217 | 17 | Q8TP72 | Q8tp72 methanosarc |

ALIGNMENTS

RESULT 1

| | | | | | |
|--------|--|---|--------------|------|----------|
| Q51839 | ID | Q51839 | PRELIMINARY; | PRT; | 1706 AA. |
| AC | Q51839 | Q51840; | | | |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Created) | | | |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Last sequence update) | | | |
| DT | 01-OCT-2003 | (TREMBLrel. 25, Last annotation update) | | | |
| DE | Arginine-specific thiol protease precursor. | | | | |
| GN | PRTR. | | | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales; | | | | |
| OC | Porphyromonadaceae; Porphyromonas. | | | | |
| OX | NCBI_TaxID=837; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=W50; | | | | |
| RX | MEDLINE=95160709; PubMed=7857299; | | | | |
| RA | Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N., | | | | |
| RA | Reynolds E.C.; | | | | |
| RT | "Complete nucleotide sequence of a gene prTR of Porphyromonas | | | | |
| RT | gingivalis W50 encoding a 132 kDa protein that contains an arginine- | | | | |
| RT | specific thiol endopeptidase domain and a haemagglutinin domain."; | | | | |
| RL | Biochem. Biophys. Res. Commun. 207:424-431(1995). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=W50; | | | | |
| RX | MEDLINE=96311339; PubMed=8713096; | | | | |
| RA | Slakeski N., Cleal S.M., Reynolds E.C.; | | | | |
| RT | "Characterization of a Porphyromonas gingivalis gene prTR that encodes | | | | |
| RT | an arginine-specific thiol proteinase and multiple adhesins."; | | | | |
| RL | Biochem. Biophys. Res. Commun. 224:605-610(1996). | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=W50; | | | | |
| RA | Reynolds E.; | | | | |
| RL | Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases. | | | | |
| RN | [4] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=W50; | | | | |

RA Slakeski N.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 126341; AAC18876.1; --
 DR HSSP; P95493; 1CVR.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25_1.
 DR Pfam; PF03785; Peptidase_C25_C_1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW PROTEASE; Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 228 1706 ARGININE-SPECIFIC THIOL PROTEASE.
 SQ SEQUENCE 1706 AA; 185626 MW; E8BDF07C9813B844 CRC64;
 Query Match 36.0%; Score 951.5; DB 2; Length 1706;
 Best Local Similarity 31.9%; Pred. No. 2.7e-50;
 Matches 237; Conservative 74; Mismatches 156; Indels 275; Gaps 20;
 QY 22 TAAAGG-----GKTPAPSVTHQAVQK--GIR---TSKVLDLRDIPAGMARILIEAH 68
 DB 671 TATTGGQKVTWKWDAPSTKTNNATTARSVDGIRELVLLSVSDAPELRSQGAIEVLEAH 730
 QY 69 DWEDGTGYQMLDADHNOYASIPPE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
 DB 731 DVNNDGSGYQILLDADHDQYGVIPSDHTLW--PNCSPANLFAFEYTVPENADPSCSP 789
 QY 127 TNFVLDTGASADIPAGTYDYVIINPGLI--IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185
 DB 790 TMIMWDGTASVNIPAGTYDYFIAAFOANAKIWIAGGPTKEDDYVFEAGKTYHFLMKWG 849
 QY 186 PGDAASV----- 193
 DB 850 SGGGTETLTISEGGGSDTYTVYVRDGTCKIKEGLTATTFREDGATGNHCEYVEVKTAGVS 909
 QY 194 -----VTGEGNEFAPQVQLQMSVGSQVTLTWQAPAS-----DKRTYVLNE 235
 DB 910 PKVKDQVTVGSENEFAPQVQLTGSVAGQKVTWKWDAPNGTPNPENPNPNTGTTLSE 969
 QY 236 SFDQTQLPNGWMTIDADGDHNLSTINVTATHTGAMFESKSWTASGGAIDLSPDN 295
 DB 970 SFE--NGIPASWKTIDADGDHGW--KPGNAPGIAGNSNGCVYSSEF--GLGGIGV--LTPDN 1025
 QY 296 YLVTPKVTVPNGKLSYVWSSQ--VFWTNEHYGVFLSTTNGEAAFTIKLLEETL---GSD 351
 DB 1026 YLITPALDLPNGKGLTFWVCAQDANYASEHYAVVASSTGNDASFTNALLETITAKGVR 1085
 QY 352 KPAPW----- 356
 DB 1086 SPEAMRGRIQWTWRQKTVLDLPAGTKYVAFRHFQSDTFYIDLDEVEIKANGKRAFDTET 1145
 QY 357 -----NLVKS----- 361
 DB 1146 ESSTHGEAPAEWTIDADGCGQWLCLSGGQDMLTAHGNTNVVSSFSWNGNALNPNYL 1205
 QY 362 ----- 361
 DB 1206 ISKDVTKATKVKYVAVNDGPFQDHYAVIMSKTGTNAGDFTVVFETPENGKGGARFGL 1265
 QY 362 -----EGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDV--AVSGEGSND 415
 DB 1266 STEADGAKPQSWIERTVDLP--AGTK--YVAFRHYNCSDLIYILLDDITQFTMGSGPTPTD 1322
 QY 416 YTYTVYRDNVVIAQNLAAATFENQENAVPQGNCEYVEVKTAGVSKCKVQTVVSGSNEFA 475
 DB 127 TNFVLDTGASADIPAGTYDYVIINPGLI--IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185

1323 YTYTVYRDGTCKIKEGLTATTFREDGATGNHCEYVEVKTAGVSKCKVQTVV--NSTQFN 1381

476 HVONLTGSAVGQKVTWKWDAPN 497

1382 PVKNLKAQPDGDDGVVLKWEAPS 1403

RESULT 2

Q51838 PRELIMINARY; PRT; 1706 AA.

AC Q51838;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Protease precursor.
 GN PRPRL.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RX MEDLINE=96071894; PubMed=7591131;
 RA Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;
 RT "Characterization, genetic analysis, and expression of a protease
 antigen (PrpRL) of Porphyromonas gingivalis W50.";
 RL Infect. Immun. 63:4744-4754 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;
 RT "The prpRL and the prx2 arginine-specific protease genes of
 Porphyromonas gingivalis W50 produce five biochemically distinct
 enzymes.";
 RL Mol. Microbiol. 23:0-0 (1997).
 DR EMBL; X82680; CAA57997.1; --
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25_001;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25_1.
 DR Pfam; PF03785; Peptidase_C25_C_1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW SIGNAL.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 228 719 ALPHA-PROTEASE.
 FT CHAIN 720 1262 BETA-ADHESIN.
 SQ SEQUENCE 1706 AA; 185705 MW; 0E56DCD87FDA8CDD CRC64;
 Query Match 36.0%; Score 950.5; DB 2; Length 1706;
 Best Local Similarity 31.8%; Pred. No. 3.1e-50;
 Matches 236; Conservative 77; Mismatches 154; Indels 275; Gaps 20;
 QY 22 TAAAGG-----GKTPAPSVTHQAVQK--GIR---TSKVLDLRDIPAGMARILIEAH 68
 DB 671 TATTGGQKVTWKWDAPSTKTNNATTARSVDGIRELVLLSVSDAPELRSQGAIEVLEAH 730
 QY 69 DWEDGTGYQMLDADHNOYASIPPE--SFWFANGTIPAGLYDPFEYKVPVNADASFSP 126
 DB 731 DVNNDGSGYQILLDADHDQYGVIPSDHTLW--PNCSPANLFAFEYTVPENADPSCSP 789
 QY 127 TNFVLDTGASADIPAGTYDYVIINPGLI--IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185

Db 790 TNMIMDGTASVNI PAGTYDFAIAAPQANAKIWIAGQGPTEKEDYVFEAGKKYHFLMKMG 849
 QY 186 PGDAASVV----- 193
 Db 850 SGDGTGLTISEGGSDYTYVVRDGTKEGLTATTPREDGVAAGNHEVCVEVKTAGVS 909
 QY 194 -----VTGEGNNEFAPVQNLQMSVGGQTVLTWQAPAS-----DKRTYVLNE 235
 Db 910 PKVKCDVTVEGSNEFAPVQNLQMSVGGQTVLTWQAPAS-----DKRTYVLNE 235
 QY 236 SFDQTQTPNGMTIDADGDGHNWLTSTINVTATHTGCMFMSKSWTASGAKIDLSPDN 295
 Db 970 SFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSSEF-GLGGIGV-LTPDN 1025
 QY 296 YLVTPKVTVPNGKLSYVWSQ-VPTWNEHYGVFLSTGTGNEAANFTIKLEETLGS----- 350
 Db 1026 YLIITPALDLPNGKLTFWVCAQDANYASEHYAVYASSGTGNDASFTNALLEETITAKVR 1085
 QY 351 -----DKPA----- 354
 Db 1086 SPEAIRGIQSTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGRADFTET 1145
 QY 355 -----PMLNV 359
 Db 1146 ESSTHGEATAEWITIDADGDGQWMLCLSSGQDLWLTAGHTNNVSSFSWNGMALNPDNYL 1205
 QY 360 ----- 359
 Db 1206 ISKDVTKATKYYVAVNDGPGDHYAVMISKGTGNTAGDFTVVFETPENGKNGARFGL 1265
 QY 360 --KSEGVKLPAQYQERTIDISAYAGQVYLAFRHFNSTGIFRLYLDV--AVSGEGSSND 415
 Db 1266 STEADGAKPQSWIERTVDLP-AGTK-YVAFRHYNGSDLYILLDDITQFTMGSSPTPTD 1322
 QY 416 YTVTVRDNVVIQNLAAATFNQENVAPQVNYCQVVKYTAGVSPKCKDVTVEGSNEFA 475
 Db 1323 YTVTVRDTGKIKGLTETTFEEDGVAAGNHEVCVEVKYTAGVSPKCKVNVTV-NSTQFN 1381
 QY 476 HVQNLTGSAVGQKVTWKWDAPN 497
 Db 1382 PVNKLQAQPDGVDVLKWEAPS 1403

RESULT 3

Q51816 PRELIMINARY; PRT; 1704 AA.
 AC Q51816; (1)
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arg-gingipain-1 proteinase.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95138080; PubMed=7836351;
 RA Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
 RA Travis J., Barr P.J.;
 RT "Molecular cloning and structural characterization of the Arg-
 RT proteinase-adhesin polypeptide."
 RL J. Biol. Chem. 270:1007-1010(1995).
 DR EMBL; U15282; AAA69539.1; --
 DR PIR; A55426; A55426.
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.003; --
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.

DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR InterPro; IPR001110; IG-like.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase C25_C.
 DR Pfam; PF01364; Peptidase C25; I.
 DR Pfam; PF03785; Peptidase C25_C; 1.
 DR PROSITE; PS00699; DNA_LIGASE_A1; 1.
 FT CHAIN 228 719 MATURE 50-KDA CYSTEINE PROTEINASE
 FT GINGIPAIN.
 SQ SEQUENCE 1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;
 Query Match 35.8%; Score 944.5; DB 2; Length 1704;
 Best Local Similarity 31.9%; Pred. No. 7.4e-50;
 Matches 235; Conservative 75; Mismatches 160; Indels 287; Gaps 20;
 QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVQLRDPDPAGMARIIIEAH 68
 Db 671 TATTQGGKVTWKWDAPSTKTNATNTARSVDGIRELVLLSVSDAPELRGQAEIVLEAH 730
 QY 69 DVWEDGTGYQMLMDADHNOYGASIPER--SFWFANGTIPAGLYDPFPEYKVPVNDASFS 126
 Db 731 DVWMDGSGYQILLDADHDQYGVIPSDHTLM-PNCVSPANLPAPFEYTPENADPSCSP 789
 QY 127 TNFVLDTGASADIPAGTYDYVIINPNFI--IYIVGEGVSKNDYVVEAGKYHTFTVORQ 185
 Db 790 TNMIMDGTASVNI PAGTYDFAIAAPQANAKIWIAGQGPTEKEDYVFEAGKKYHFLMKMG 849
 QY 186 PGDAASVV----- 193
 Db 850 SGDTELTISEGGSDYTYVVRDGTKEGLTATTPREDGVAAGNHEVCVEVKTAGVS 909
 QY 194 -----VTGEGNNEFAPVQNLQMSVGGQTVLTWQAPAS-----DKRTYVLNESEF 237
 Db 910 PKVKCDVTVEGSNEFAPVQNLQMSVGGQTVLTWQAPAS-----DKRTYVLNESEF 237
 QY 238 DTQTLNGMTIDADGDGHNWLTSTINVTATHTGCMFMSKSWTASGAKIDLSPDNYL 297
 Db 970 E-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVYSSEF-GLGGIGV-LTPDN 1025
 QY 298 YLVTPKVTVPNGKLSYVWSQ-VPTWNEHYGVFLSTGTGNEAANFTIKLEETLGS----- 350
 Db 1026 YLIITPALDLPNGKLTFWVCAQDANYASEHYAVYASSGTGNDASFTNALLEETITAKVRSP 1085
 QY 351 -----DKPAPMLNV-----KSEGVKL----- 366
 Db 1086 EAIRGRIQSTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGRADFTETTES 1145
 QY 367 ---PAPYQERTID-----LSAYAGQV----- 385
 Db 1146 STHGEATAEWITIDADGDGQWMLCLSSGQDLWLTAGHTNNVAVSFWNGMALNPDNYLIS 1205
 QY 386 ----- 385
 Db 1206 KDVTKATKYYVAVNDGPGDHYAVMISKGTGNTAGDFTVVFETPENGKNGARFGLST 1265
 QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTYV 420
 Db 1266 EANGAKPQSWIERTVDLPAGTKYVAFRHYNGSDLYILLDDITQFTMGSSPTPTDYYTV 1325
 QY 421 YRDNVVIQNLAAATFNQENVAPQVNYCQVVKYTAGVSPKCKDVTVEGSNEFAHVQNL 480
 Db 1326 YRDTGKIKGLTETTFEEDGVAAGNHEVCVEVKYTAGVSPKCKVNVTV-NPTQFNPKNL 1384
 QY 481 TGSVAGQKVTWKWDAPN 497
 Db 1385 KAQPDGVDVLKWEAPS 1401
 ID Q9R9B7 PRELIMINARY; PRT; 1687 AA.
 RESULT 4
 Q9R9B7

AC Q99B7;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hemagglutinin/protease.
 GN HAGE.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RA Han N., Dong H., Progluske-Fox A.;
 RT "Cloning and characterization of hage from P. gingivalis 381."
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF026946; AAD01810.1;
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; I.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64;

 Query Match 35.7%; Score 943.5; DB 2; Length 1687;
 Best Local Similarity 31.8%; Pred. No. 8.4e-50;
 Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

 QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR--TSKYKDLRDPDPAGMARILIEAH 68
 DB 652 TATTGGKVTWKWAPSTKTWATNTAKSVGIRLVLSVSDAPELRSGQAEIVLEAH 711

 QY 69 DVWEDGTGYQLWADHNNQYASIPEE--SPWFANGTIPAGLYDPFVKYKVPVNDAGFSP 126
 DB 712 DVWNGSGYQLLDADHDQYGVIPSDTHILW--FNCSPANLPAPFYTPENADPSCSP 770

 QY 127 TNFVLDGTASADIPAGTYDYVIINPNFGI--IYIVGEGVSKNDYVVEAGKTYHFTVQROG 185
 DB 771 TNMIMDGTASVNIIPAGTYDYFAIAAPQANAKIWIAGOGPTKEDDYVFEAGKKYHFMKMG 830

 QY 186 PGDAASVV-----193

 DB 831 SGDGTETLTISEGGSDYTYVYRDGTIKLKEGLTATTEEGVATGNHEYCVVEKYTAGVS 890

 QY 194 -----VTGEGGNFAPVQMLWSVSGTQVTLTWQAPAS-----DKRTYVINE 235
 DB 891 PKVKCKDVTEGSENEFAPVQNLGTSAGVQKVTWKWDAPNGTPNPNPNPNFPGTTLSE 950

 QY 236 SEDTQTLPGNWMTLDADGDGHNWLTINVTNATHTGCGAMFWSKWTASGAKIDISPDN 295
 DB 951 SFE--NGIPASWKTLDADGDGHW--KPGNAPFGIAGYNSNGCVYSEF--GLGGIGV-LTPDN 1066

 QY 296 YLTPPKVTVPENGKLSYVWSQ--VPWTNEHYGVFLSTTGNEAANFTIKLEETLGS----- 350
 DB 1007 YLITPALDLNGKLTFWCAQANYASEHYAVYASSTGNDASNTNALLEETITAKVR 1066

 QY 351 -----DKPAPMNLV-----KSEGVKL-----366
 DB 1067 SPEAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSDMFYIDLDEVEIKANGKADTETF 1126

 QY 367 -----PAPQERTID-----LSAYAGQOV-----385

DB 1127 ESSTHGEAPAEWTTIDADGCGQWMLCUGSGOLDMLTAHGGTNVVSFSGWGMALPNPNYL 1186
 QY 386 -----385

 DB 1187 ISKDVGTGATKVKYVAVNDGFPDHYAVMISKGTNAGDFTVVFETPNGINKGAREGL 1246
 QY 386 -----YLAFRHNTGTIFRLYLDV--AVSGEGSSNDYTY 418
 DB 1247 STEANGAKPQSWIERTVLDLPAGTKYVAFRHYNGSDLYILLDDIQFTMGSPPTDYTY 1306
 QY 419 TVYRDNVVIAQNLAAATTFNENAVPGOVNYCVVEKYTAGVSPKCKOVTVSGSSEFAHQV 478
 DB 1307 TVYRDGTIKLKEGLTETFEEDGVTAGNHEYCVVEKYTAGVSPKCEVNVTTI--NPTQFNPK 1365
 QY 479 NLGSAVGQKVTWKWDAPN 497
 DB 1366 NLKAQPDGDDVVLKWEAPS 1384

 RESULT 5
 ID P72194 PRELIMINARY; PRT; 1723 AA.
 AC P72194;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Lys-gingipain.
 OS KGP.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RX MEDLINE=97044756; PubMed=8898827;
 RA Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
 RT "Cloning and sequencing of the gene encoding a novel lysine-specific
 cysteine proteinase (lys-gingipain) in Porphyromonas gingivalis:
 structural relationship with the arginine-specific cysteine proteinase
 (Arg-gingipain).";
 RT J. Biochem. 120:398-406(1996).
 RL EMBL; D83258; BAA11870.1;
 DR MEROPS; C25.002;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; I.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1723 AA; 187261 MW; 5628963D251493EB CRC64;

Query Match 31.2%; Score 824; DB 2; Length 1723;
 Best Local Similarity 29.8%; Pred. No. 2.3e-42;
 Matches 220; Conservative 75; Mismatches 170; Indels 274; Gaps 21;

 QY 22 TAAAGGPKT-----APSVTHQAVQKIRTSKYKDLRD-----PIPAGMARILIEA 67
 DB 693 TATTGGKVTWKWDAPS-----AKKAEASREVKRGDGLFVTIEPANDVRANEAKVLLAA 747

 QY 68 HDVWEDGTGYQLWADHNNQYASIPESFWEANGTIPAGLYDP--FEYKVPVNDAGFSP 126
 DB 748 DNVWNGTGYQLLDADHDHNTFGSVIPATGFLP--TGTASSNLNYSANFEYLIPANAPVVT 806

 QY 127 TNFVLDGTASADIPAGTYDYVIINPNP--GIIVYVGEV--VSKGNDYVVEAGKTYHFTV 181

```

Db 807 QNIIVTQGGVVIQGGVYDYCIITNPEPASGKMWIAGDGNQPARYDDFTFEAGKYYTFTM 866
Qy 182 QROGPGDAASVV----- 193
Db 867 RRAGMGDTMEVEDDSSPASYTYVYRDGKIQELGTATTEEDGVAAGNHEYCVVEKYT 926
Qy 194 -----VTGEGNEFAPVQNLQMSVSGQTVTLTWQAPAS-----DKRTYVINE 235
Db 927 AGVSPKVKCDVTVEGSENEFAPVQNLTGSAVGQKVLKWDANGTNPENPNTGTTILSE 986
Qy 236 SFDQTLPNGWTMDADGCHNWLSTINVTNTHGDMGAMFSKSWTASGAKIDLSPDN 295
Db 987 SFE-NGIPASWKTTDADGCHGW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPDN 1042
Qy 296 YLTPKVTVPENGKLSYVWSSQ-VPTNHEHYGVFLSTTGNAAFTIKLEETLGS----- 350
Db 1043 YLITPDLDFNGKLTFWCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1102
Qy 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 1103 SPEAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFGQTMDFYIDLDEVEIKANGKRAFDTET 1162
Qy 367 -----PAPYQERTID-----LSVAGQV----- 385
Db 1163 ESTHGEAPAEWTTIDADGQDQWLCLSSGQDLWLTAGHTNVVASFWSWNGMALNPDNYL 1222
Qy 386 ----- 385
Db 1223 ISKDVGTATKVKYKYAVNDGFGPDHYAVMISKTGTNAGDFTVVEETPNGINKGARFGL 1282
Qy 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTY 418
Db 1283 STEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQTMTGSPPTTLY 1342
Qy 419 TVYRDNVVIAQNLAAFTNQNENAPQGYNYCVVEKYTAGVSPKVKCDVTVEGSENEFAHVQ 478
Db 1343 TVYRDGKIKKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKVCNVVTI-NPTQFNPK 1401
Qy 479 NLTGSAVGQKVLKWDAPN 497
Db 1402 NLKAQPDGGDVVLKWEAPS 1420

```

RESULT 6

```

P72197 ID P72197 PRELIMINARY; PRT; 1723 AA.
AC P72197
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lys-gingipain.
GN KGP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RS SEQENCE FROM N.A.
RA Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT "Molecular cloning and characterization of Porphyromonas gingivalis
RT Lys-gingipain."
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U54691; AAA99810.1; --
DR MEROPS; C25.002; --
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

```

```

DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005336; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1723 AA; 186831 MW; 4508A7E50197CEBD CRC64;

Query Match 31.0%; Score 818; DB 2; Length 1723;
Best Local Similarity 25.4%; Pred. No. 5.4e-42;
Matches 218; Conservative 78; Mismatches 166; Indels 280; Gaps 21;

Qy 22 TAAAGGPKT-----APSVTHQAVQKGIKTSKVKDLRD-----PIPAGMARILEA 67
Db 693 TATTQGGKVLKWDAPS-----AKKAGSREVKKRIGDGLFVTIEBPANDVRANEAKVILAA 747
Qy 68 HDVWEDGTGYQMLDADHNOYGSIPESFWEFANGTIPAGLYDP-FEYKVPVNADASFSP 126
Db 748 DNVWGDNTGYQLLDADHNTFGSVIPATGLF-TGTASSNLYSANFEYLIPANADPVVTT 806
Qy 127 TNFVLDGTASADIPAGYDYVINPNP--GLIYIVGEG--VSKGNDYVVEAGKTYHTV 181
Db 807 QNIIVTQGGVVIQGGVYDYCIITNPEPASGKMWIAGDGNQPARYDDFTFEAGKYYTFTM 866
Qy 182 QROGPGDAASVV----- 193
Db 867 RRAGMGDTMEVEDDSSPASYTYVYRDGKIQELGTATTEEDGVAAGNHEYCVVEKYT 926
Qy 194 -----VTGEGNEFAPVQNLQMSVSGQTVTLTWQAPAS-----DKRTYVINE 235
Db 927 AGVSPKVKCDVTVEGSENEFAPVQNLTGSAVGQKVLKWDANGTNPENPNTGTTILSE 986
Qy 236 SFDQTLPNGWTMDADGCHNWLSTINVTNTHGDMGAMFSKSWTASGAKIDLSPDN 295
Db 987 SFE-NGIPASWKTTDADGCHGW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPDN 1042
Qy 296 YLTPKVTVPENGKLSYVWSSQ-VPTNHEHYGVFLSTTGNAAFTIKLEETLGS----- 350
Db 1043 YLITPDLDFNGKLTFWCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVR 1102
Qy 351 -----DKPA----- 354
Db 1103 SPEAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFGQTMDFYIDLDEVEIKANGKRAFDTET 1162
Qy 355 -----PMLV 359
Db 1163 ESTHGEAPAEWTTIDADGQDQWLCLSSGQDLWLTAGHTNVVASFWSWNGMALNPDNYL 1222
Qy 360 ----- 359
Db 1223 ISKDVGTATKVKYKYAVNDGFGPDHYAVMISKTGTNAGDFTVVEETPNGINKGARFGL 1282
Qy 360 --KSEGVKLPAPOERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDV--AVSGEGSSND 415
Db 1283 STEANGAKPQSVMIERTVDLP--AGTK-VYAFRHYNCSDLNYILLDDIQTMTGSPPTTD 1339
Qy 416 YTVYRDNVVIAQNLAAFTNQNENAPQGYNYCVVEKYTAGVSPKVKCDVTVEGSENEFA 475
Db 1340 YTVYRDGKIKKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKVCNVVTI-NPTQFN 1398
Qy 476 HVQNLTGSAVGQKVLKWDAPN 497
Db 1399 PVKNLKAQPDGGDVVLKWEAPS 1420

RESULT 7
P96967 ID P96967 PRELIMINARY; PRT; 1358 AA.
AC P96967;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin.

```



```

GN HAGD.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA Han N., Lepine G., Whitlock J., Wojciechowski L., Progulski-Fox A.;
RT "Cloning, sequencing and characterization of hcp, a member of the
RT HAREP multigene family in Porphyromonas gingivalis.";
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ Databases.
DR EMBL; U68468; AAB49691.1; -.
DR GO:0005524; F:ATP binding; IEA.
DR GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO:0006310; P:DNA recombination; IEA.
DR GO:0006281; P:DNA repair; IEA.
DR GO:0006260; P:DNA replication; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR001769; Peptidase_C25.
DR Pfam; PF01364; Peptidase_C25; I.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

Query Match 30.9%; Score 816; DB 2; Length 1358;
Best Local Similarity 29.6%; Pred. No. 5.1e-42;
Matches 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKTSKVKDLRD-----PIPGMARIIIEA 67
DB 328 TATTGGQKVTWKWDAPS-----AKKAESREVKRIGDGLFVTIEPANDVRANEAKVVLAA 382

QY 68 HDVWEDGTGYQMLDADHNOYGASIPERSFWFANGTIPAGLYDP-FEYKVPWNADASFSP 126
DB 383 DNVMGDNTGYQLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 441

QY 127 TNFVLDTGASADIPAGTYDYVIINPNP-GIIYIVGEG---VSKGNDYVVEAGKYHTFT 181
DB 442 QNIIVTGGVEVPIPGVYDYCIITNPEPASGKMWIAGDGNOPARYDDFTFEAGKKYFTM 501

QY 182 QRQPGGDAASVV----- 193
DB 502 RRAGMGDGTDEVEDDSDPSYTYTYVRDGTKIKEGLTATTFEEDGVAAGNHEYCVVKYT 561

QY 194 -----VTGEGNEFAPQNIQWVSQGTVTLTWQAPAS-----DKRTYVLNE 235
DB 562 AGVSPKVKCDVTVEGSENFAPQNLITGSAVGQKVTWKWDAPNGTNPENPNPGTTILSE 621

QY 236 SFDTQTLNGMTIDADGDHNLSTINVTATHTGGMFSGKSWTASGAKIDLSPDN 295
DB 622 SFE-NGIPASWKTIDADGDHGW-KPGNAPGIAGYNSGCVYSESF-GLGGIGV-LTPDN 677

QY 296 YLTPKVTVPENGKLSYVSSQ-VPWNTNEHGVFLSTTGNAAFTIKLLEETLGS---- 350
DB 678 YLITPALDANGKKLITFWVCAQDANYASEHAYAVASSTGNDSNFTNALLRETTAKVGR 737

QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
DB 738 SPEAIRGIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFVIDLDEVEIKANGKRADFTTF 797

QY 367 -----PAPYQERTID-----LSAVAGQOV----- 385
DB 798 ESSTHGEAPAEWTIIDAGDGDWMLCLSSGQDMLTATHGNTVWVASFWSNGMALNPNDYL 857

QY 386 ----- 385
DB 858 LSKDVTGATKYYVAVNDGFGDHYAVMI SKTGTNAGDFTVVPETENGINKGARFGL 917

QY 386 -----YLAPRHNSTGIFRLYLLDDV--AVSGEGSNDYTY 418
DB 918 STEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNVLDDIQTMTGGSFTPTDITY 977

QY 419 TVYRDNVVIAQNLAAATTENQENVAPGOXNYCVVEKYTAGVSPKCKDVTVEGSEFAHVQ 478
DB 978 TVYRDGTKIKEGLTTFTEEDGVAAGNHEYCVVEKYTAGVSPKCVNVTI-NPTQFNPKV 1036

QY 479 NLTGSAVGQKVTWKWDAPN 497
DB 1037 NLKAQPDGDDGVVLKWEAPS 1055

RESULT 8
Q51817
ID Q51817 PRELIMINARY; PRT; 1732 AA.
AC Q51817;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Porphyropain.
GN PRTp.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W12;
RX MEDLINE=96213011; PubMed=8631659;
RA Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA Progulski-Fox A., Lantz M.S.;
RT "Analysis of the prcp gene encoding porphyropain, a cysteine proteinase
RT of Porphyromonas gingivalis.";
RL J. Bacteriol. 178:2734-2741 (1996).
DR EMBL; U42210; AAB06565.1; -.
DR PIR; T30836; T30836.
DR MEROPS; C25_002;
DR GO:0005524; F:ATP binding; IEA.
DR GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO:0006310; P:DNA recombination; IEA.
DR GO:0006281; P:DNA repair; IEA.
DR GO:0006260; P:DNA replication; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; I.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCAE4 CRC64;

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1.2e-41;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKTSKVKDLRD-----PIPGMARIIIEA 67
DB 693 TATTGGQKVTWKWDAPS-----AKKAESREVKRIGDGLFVTIEPANDVRANEAKVVLAA 747

QY 68 HDVWEDGTGYQMLDADHNOYGASIPERSFWFANGTIPAGLYDP-FEYKVPWNADASFSP 126
DB 748 DNVMGDNTGYQLLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806

QY 127 TNFVLDTGASADIPAGTYDYVIINPNP-GIIYIVGEG---VSKGNDYVVEAGKYHTFT 181
DB 807 QNIIVTGGVEVPIPGVYDYCIITNPEPASGKMWIAGDGNOPARYDDFTFEAGKKYFTM 866

QY 182 QRQPGGDAASVV----- 193
DB 867 RRAGMGDGTDEVEDDSDPSYTYTYVRDGTKIKEGLTATTFEEDGVAAGNHEYCVVKYT 926

```

```

QY 194 -----VTGCGNEFAPQNLQWSVSGQTVTLTWQAP-----ASDKRTYVLN 234
DQ 927 AGVSPKVKCDVTVEGSENEFAPQNLQWSVSGQTVTLTWQAP-----ASDKRTYVLN 234
QY 235 ESPTDTQLPNGWTMDADGDGHNWLTSTINVTNTATHTGDMGAMFSKSWTASGAKIDLSPD 294
DQ 987 ESFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVSESF-GLGGIGV-LTPD 1042
QY 295 NYLVTPKVTVPENGKLSYVSSQ-VPTWNEHYGVFLSTTGNANFTIKLLEETLS--- 350
DQ 1043 NYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASFTNALLEETITAKGV 1102
QY 351 -----DKPAPMNLV-----KSEGVKLV----- 366
DQ 1103 RSPKAIIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLDEVEIKANGRADTET 1162
QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
DQ 1163 FESSTHGEAPAEWTTIDADGDGQWCLSSGQLDLTAHGSNNVSVFWSNGMALNPNDY 1222
QY 386 ----- 385
DQ 1223 LISKQVGTATKYYAVNDGPPGDHYAVMISKTGTNAGDFTVVEETPNKNGGARFG 1282
QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
DQ 1283 LSTEANGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDNLVILLDDIQFTMGSPPTDYT 1342
QY 418 YTVYRDNVVIQNLAAATFNQENVAPQYNYCEVVKYTAGVSPKVKCDVTVEGSENEFAHV 477
DQ 1343 YTVYRDTGTLKEGLTETTFEEDGVTATGNHEYCEVVKYTAGVSPKVKCDVTV-NSTQFNVP 1401
QY 478 QNLTSAGV--GQKVTILKWDAP 496
DQ 1402 QNLTAQAPNSMDAILKNAP 1422

RESULT 9
O52050
ID O52050 PRELIMINARY; PRT; 1732 AA.
AC O52050;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lysine specific cysteine protease.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=98298016; PubMed=9632563;
RA Lewis J.P., Macrina F.L.;
RT "IS195, an insertion sequence-like element associated with protease
genes in Porphyromonas gingivalis."
RL Infect. Immun. 66:3035-3042(1998).
DR EMBL; AF017059; AAC26523.1;
DR MEROPS; C25.002;
DR GO; GO:0005524; F-ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase C25.
DR Pfam; PF01364; Peptidase C25; I_1.
DR Pfam; PF013785; Peptidase C25; C_1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.

```

```

KW Protease.
SQ SEQUENCE 1732 AA; 187931 MW; B2337463D5CBSEAS CRC64;

Query Match 30.6%; Score 807.5; DB 2; Length 1732;
Best local Similarity 30.0%; Pred. No. 2.4e-41;
Matches 22; Conservative 69; Mismatches 173; Indels 277; Gaps 22;

QY 22 TAAAGGPKT-----APSVTHQAVQKGIKRTSKVDLRD-----PIFAGMARIILEA 67
DQ 693 TATTQGGKVTLKWEAPS-----AKKAEGRSEVRKIGDGLFVFTIEPANDVRANEAKVVLAA 747
QY 68 HDVWEDGTGQMLWDADHNOYGASIPESFWEANGTIPAGLYDP-FEYKVPVNDASFS 126
DQ 748 DNVWGDNTGQFLLDADHNTFGSVIPATGPIF-TGTASSNLIYANFYLVPANAPVVT 806
QY 127 TNFVLDGASADIPAGTYDYIINPNP--GIYIVGEG--VSKGNDYVVEAGTYHTFT 181
DQ 807 QNIIVTQGGEWIPGGVYDYICITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYFTM 866
QY 182 QROGPGDAASV----- 193
DQ 867 RRAGMGDTGDMVEDDPSASYTYTVYRDGTIKI EGLTATTTFEEDGVAAGNHEYCEVVKY 926
QY 194 -----VTGCGNEFAPQNLQWSVSGQTVTLTWQAP-----ASDKRTYVLN 234
DQ 927 AGVSPKVKCDVTVEGSENEFAPQNLQWSVSGQTVTLTWQAP-----ASDKRTYVLN 234
QY 235 ESPTDTQLPNGWTMDADGDGHNWLTSTINVTNTATHTGDMGAMFSKSWTASGAKIDLSPD 294
DQ 987 ESFE-NGIPASWKTIDADGDGHW-KPGNAPGIAGYNSGCVSESF-GLGGIGV-LTPD 1042
QY 295 NYLVTPKVTVPENGKLSYVSSQ-VPTWNEHYGVFLSTTGNANFTIKLLEETLS--- 350
DQ 1043 NYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASFTNALLEETITAKGV 1102
QY 351 -----DKPAPMNLV-----KSEGVKLV----- 366
DQ 1103 RSPKAIIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLDEVEIKANGRADTET 1162
QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
DQ 1163 FESSTHGEAPAEWTTIDADGDGQWCLSSGQLDLTAHGSNNVSVFWSNGMALNPNDY 1222
QY 386 ----- 385
DQ 1223 LISKQVGTATKYYAVNDGPPGDHYAVMISKTGTNAGDFTVVEETPNKNGGARFG 1282
QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
DQ 1283 LSTEANGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDNLVILLDDIQFTMGSPPTDYT 1342
QY 418 YTVYRDNVVIQNLAAATFNQENVAPQYNYCEVVKYTAGVSPKVKCDVTVEGSENEFAHV 477
DQ 1343 YTVYRDTGTLKEGLTETTFEEDGVTATGNHEYCEVVKYTAGVSPKVKCDVTV-NSTQFNVP 1401
QY 478 QNLTSAGV--GQKVTILKWDAP 496
DQ 1402 QNLTAQAPNSMDAILKNAP 1422

RESULT 10
O07442
ID O07442 PRELIMINARY; PRT; 1732 AA.
AC O07442;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lysine-specific cysteine proteinase.
GN PRK.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-W50;
RX MEDLINE=99235907; PubMed=10219167;
RA Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RT a lysine-specific cysteine proteinase and three sequence-related
RT adhesins."
RL Oral Microbiol. Immunol. 14:92-97 (1999).
DR EMBL; U75366; AAB60809.1; -.
DR MEROPS; C25.002; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR Pfam; PF01364; Peptidase_C25; I.
DR Pfam; PF03785; Peptidase_C25; I.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR SEQUENCE 1732 AA; 187914 MW; 45DSB91377391703 CRC64;

Query Match 30.5%; Score 806.5; DB 2; Length 1732;
Best Local Similarity 29.7%; Pred. No. 2.8e-41;
Matches 220; Conservative 72; Mismatches 172; Indels 277; Gaps 22;

QY 22 TAAAGGGPKT-----APSVTHQAVQKIRTSKVKDLRD-----PIPGMARIILEA 67
DQ 693 TATTQGGKVTWKWEAPS-----AKAEGSREVKRGDGLFVTIEPANDVRANEAKVLLAA 747
QY 68 HDVWEDGTGYQMLDADHNOYCASIPESFWFANGTIPAGLYDP-FEYKVPVNADASFSP 126
DQ 748 DNVWGDNTGYQFLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLIPANADPVVTT 806
QY 127 TNFVLDTGASADIPAGTYDYVINPNP--GIYIVGEG--VSKGNDYVVEAGKTYHFTV 181
DQ 807 QNIIVTGGGEVVPFGVYDYCIINPEPASGKMWIAGDGNQPARVDDFTFEAGKKYTFM 866
QY 182 QRQGGDAASVV-----DKPAPNLY-----KSGVKL----- 193
DQ 867 RRAGMGDGTDMVEDDPSASYTYTVYRDGTKIKEGLTATTFTFEEDGVAAGNHCYVEVKYT 926
QY 194 -----VTGEGNEFAPVQNLQWSVSGQVTLTWOAP-----ASDKRYTVLN 234
DQ 927 AGVSPKVKCKVTVGSENFAPVQNLQWSVSGQVTLTWOAP-----PIPGMARIILEA 67
QY 235 ESFDTQTLPLNGWTMIDADGDGHNWLSLTINVTATHTGDMFESKSWTASGAKIDLSPD 294
DQ 987 ESFE-NGIPASKKITDADGDGHW-KPNAPGIAGYNSGCVYSSEF-GLGGIGV-LTPD 1042
QY 295 NYLTPKVTVPENGLKSWVSSQ-VFWTNEHYGVFLSTTGNEAANFTIKLEETLGS--- 350
DQ 1043 NYLTIPALDLPLNGGKLTFWCAQDANAYASEYAVVASTGNDASNTLALLEETITAKGV 1102
QY 351 -----DKPAPNLY-----KSGVKL----- 366
DQ 1103 RSPKAIRGRIQGWNRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGRADETET 1162
QY 367 -----PAPYQERTID-----LSAYAGQV----- 385
DQ 1163 FESSTHGAPAEWTIDADGDGQGWLCUSSQLDNLTAHGSNVVSVSWNGMALNPNDY 1222
QY 386 ----- 385
DQ 1223 LISKDVTKATKVKYVAVNDGPGPDHYAVMISKTGTNAGDFTVVFETEPNGKGAFFG 1282
QY 386 -----YLAFFHFNSTGIFRLYLDVV--AVSGEGSSNDYT 417
DQ 1283 LSTEANGAKPOSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLDDIQFTMGSGSPPTDYT 1342

```

```

Db 478 ESPE-NGIPASWKTIDADGGHGW-KPGNAPGAGYNSGCVYSSEF-GLGGIGV-LTPD 533
QY 295 NYLVTPKVTVPENGLKSYWSSQ-VPWNEHYGVFSLTTCNEAANFTIKLEETLGS--- 350
Db 534 NYLITPALDLPNGGKLTFFWCAQADANYASBHYAVYASSTGNDASNFNALLEETITAKGV 593
QY 351 -----DKPAPMLNV-----KSEGVKL----- 366
Db 594 RSEPAIRGRIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFTET 653
QY 367 -----PAPYQERTID-----LSAYAGQV----- 385
Db 654 FESSTHGEAPAEWTTIDADGGQWCLSSGQLDWTAGHTNVSFWSWNGMALNPDNY 713
QY 386 ----- 385
Db 714 LISKDVGTATKVKYVAVNDGFGPDHYAVMISKTGTNAGDFTVVFETPENGINKGARFG 773
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYT 417
Db 774 LSTEANGAKPQSVMIERTVLPAGTKYVAFRHYNCSDLNVLDDIQTWGGSPPTDYI 833
QY 418 YTVYRNVVIAQNLAAATFNQENVAPGQYNYCYEVKVTAGVSPKVCXDVTEGSENEFAHV 477
Db 834 YTVIRDTKIKELGETTFEEDGVATGNHBYCYEVKVTAGVSPKVCVNVITI-NPTQFPV 892
QY 478 QNLTGSAV--GQKVTCLKWDAP 496
Db 893 QNLTAQAPNSMDAILKKNAP 913

RESULT 12
P72196
ID Q9KIB3 PRELIMINARY; PRT; 1097 AA.
AC P72196;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TonB-linked adhesin precursor.
GN TLA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97386416; PubMed=9244265;
RA Aduse-Opoku J., Stanley J.M., Young K.A., Muir J., Rangarajan M.,
RA Curtis M.A.;
RT "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT arginine-specific protease precursor (PrpRI) which shares sequence
RT similarity to TonB-linked receptors.";
RL J. Bacteriol. 179:4778-4788(1997).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL: Y07618; CAA6897.1; -.
DR GO: GO:0019867; C:outer membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006310; P:DNA recombination; IEA.
DR GO: GO:0006281; P:DNA repair; IEA.
DR GO: GO:0006260; P:DNA replication; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR GO: GO:0009777; DNA ligase.
DR InterPro: IPR000977; DNA ligase.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF005933; TonB_dep_rec; 1.
DR PROSITE: PS00697; DNA_LIGASE_A1; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 53 POTENTIAL.
SQ SEQUENCE 1097 AA; 118731 MW; 73BBA337B421F8B9 CRC64;

```

```

Query Match 21.3%; Score 561.5; DB 2; Length 1097;
Best Local Similarity 29.8%; Pred. No. 2.4e-26;
Matches 158; Conservative 48; Mismatches 117; Indels 207; Gaps 17;

QY 165 KNDYVVEAGKT---YHFT-----VQKGPQDAASVVVTGGGNEFAFPVQNLQWSVSG 214
Db 267 KGSDDYVATGRFRFGIDFTPEWSLNYQNVFLGDAIPV---GGSNEFAFPVQNLGSSVSG 322
QY 215 QVTTLTQWAP-----ASDKRTVVLNESPTQTLPNGWTMTIDADGGHNLSTINVY 265
Db 323 QKVTCLKWDAPNGTPNPNPNPNTGTLSEF-NGIPASWKTIDADGGHGW-KPGNAP 380
QY 266 NTATHTGDGAMFSKWTASGGAKIDLSPNLYLTPKVTVPENGLKSYWSSQ-VPWNEH 324
Db 381 GIAGYNSNGCVYSSEF-GLGGIGV-LTPDNYLITPALDLPNGGKLTFFWCAQADANYASEH 438
QY 325 YGVFLSTTNEAANFTIKLEETLGS-----DKPAPMLNV----- 359
Db 439 YAVYASSTGNDASNFNALLEETITAKGVRSPKAIIRGRIQGTWRQKTVLDPAGTKYVAFR 498
QY 360 -----KSEGVKL-----PAPYQERTID----- 376
Db 499 HFQSTDMFYIDLDEVEIKANGKRAFTETFEFSTHGEAPAEWTTIDADGGQWCLSSG 558
QY 377 -----LSAYAGQV----- 385
Db 559 QLDWLTAGGSSNVVSSFSWNGMALNPDNYLISKDVGTATKVKYVAVNDGFGPDHYAVMI 618
QY 386 -----YLAFR 390
Db 619 SKTGTNAGDFTVVFETPENGINKGARFGLSTEANGAKPQSVMIERTVLPAGTKYVAFR 678
QY 391 HFNSTGIFRLYLDV--AVSGEGSSNDYTTVYRNVVIAQNLAAATFNQENVAPGQYNY 448
Db 679 HYNCSDLNVLDDIQTWGGSPPTDYITVYTRDGTIKELGETTFEEDGVATGNHBY 738
QY 449 CVEVKYTAGVSPKVCXDVTEGSENEFAHVQNLTGSAV--GQKVTCLKWDAP 496
Db 739 CVEVKYTAGVSPKVCVNVITI-NSTQFPVQNLTAQAPNSMDAILKKNAP 787

RESULT 13
Q9KIB3
ID Q9KIB3 PRELIMINARY; PRT; 312 AA.
AC Q9KIB3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF237555; AAF81413.1; -.
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match 14.4%; Score 379; DB 2; Length 312;
Best Local Similarity 37.2%; Pred. No. 8.6e-16;
Matches 94; Conservative 44; Mismatches 85; Indels 30; Gaps 9;

```

```

QY 233 LNESDQTOTLPGNWTMDADSGHNLSTINVTATHTGDMFSGKWTSGGAKIDLS 292
Db 4 LSESPES-GIPAINKTIDADGGYNWMLTN-----FTGSGLCVSSASVIGVGA-LT 55
QY 293 PDNYLAVTKVTPENG--KLSYWSOQ-VPWTNEHYGVFLSTTGNAAFTKLLBEETIG 349
Db 56 PDNYLITLDELKPTDALVEIIVWCTQDLTAPSEHYAVYSSSTGNAAADFVNLLEYETIT 115
QY 350 SDKPAPMMLV---KSEGVKLPAPYQERTIDLSAYAGQOVYLAFRHFNSTGIFRLYLDVVA 406
Db 116 AKRIQSPELIRNRTQGV---YQKVV---LNDKYVAFRHFNSTGIFRLYLDVVA 167
QY 407 V-----SEGGSSNDYTYVRNVNVAQNLAAFTTNOENAVPQYNYCVVEKYTAGV 458
Db 168 ILYTLPRAAPCPHPGGYTSYVFRDQKIASLSALAYIDTDPYGTQDYCVQVNYLQSD 227
QY 459 SPKVKCKVTVESG 471
Db 228 SYKVKCKNIVVANS 240

RESULT 14
Q9F4J0 PRELIMINARY; PRT; 925 AA.
AC Q9F4J0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DB Putative outer membrane protein PG57
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=W50;
RA Ross B.C., Crajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT "Identification of vaccine candidates from genomic analysis of
RT Porphyromonas gingivalis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY007522; AAG24228.1; -.
DR InterPro: IPR003961; FN.III.
DR SMART: SM00060; FN3; 1.
SQ SEQUENCE 925 AA; 103632 MW; 5FF2198D6914DAE8 CRC64;

Query Match
Best Local Similarity 12.4%; Score 337; DB 2; Length 925;
Matches 152; Conservative 62; Mismatches 191; Indels 194; Gaps 31;

QY 4 LNSLFSLAIVLSLCLW-----GQTAAGGQGFKTAPSVTHQAVQKGIKTSVKD 51
Db 5 MKSIVFRAFLTILLSWAITNPTAQEISGMNASCLAAPAPQDITLIESFENG----- 56
QY 52 LRDPDPAGMARILEAHVDWEDGTGYQMLWDADNINQYASIPESFWFANGT-----IPA 106
Db 57 ---FVPPNGWLEIDADA-----DGA-----TWGSPSGSF--SVP-----YCHNGLCTYSHIRS 98
QY 107 GLYDPFEEKVPVNADA-----SFQPTNFVL-----DGTASADIPAGTYDVIIN 150
Db 99 GISTAGNYLITPNLEGAKRVKYVCNQYSTNPEHYAVMVSTGTGAIEDFVLLFDDSTGK 158
QY 151 PNPGI-----IYVGEVGS-----KNDYVVEAGKTHFTVQR--QGPGDAASVVTGE 197
Db 159 PTLVWRRRIVDLPETGKYIAWRHYKYVDSHTEFLKLDVTVYRSISGPEPATDFTVINI 218

```

```

QY 198 GGNEFAPYONLQWS-----VSGQTVTLTWQAPASDK-----RTY- 231
Db 219 QQN-----VGLTWNPEDYQPEGKNEELQLSGYNIYANGTLLAQIKDVLSILEYVDSTYS 274
QY 232 -----VLNESPDQTOTLPGNWTMDADSGH 256
Db 275 LRDNPLQVEYCVTAVIDESTBSTVCGTLHYATDAILYENFENGPPVNGMLVIDADGDGF 334
QY 257 NWLSTINVT-ATHTGDMFSGKSWTSGAKIDLSPDNYLVTPTKVTVPENGK-LSVWV 314
Db 335 SWGHYLAAYDAFPNGHGHCHSLSASYVPGIG---PVTFDNYLITPKV---EGAKRVKYVW 388
QY 315 SSQ-VPWTNEHYGVFLSTTGNAAFTKLLBEETLGSDDKAPMNLVKSEGVKLPAPYQER 373
Db 389 STQANWAAEHYAWASTTGTAVGDFVI-LFEETM-TAKPT-----GAWYER 433
QY 374 TIDLSAYAGQOVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYTYVRNVV---IAQNL 431
Db 434 TINLP--EGTK-YIAWRHYNCTDIYFLKLDITVFGTPASEPEPVT---DFVSLIENNK 487
QY 432 AATTNOENAVPQYNYCVVEKYTAGVSPKVKD---VTVEGSNEFA-----HVQNL 481
Db 488 GRLKWN-----YPNGYEPDKTDDKPLQLAGYNIYANGSLLVHIQDPT 530

RESULT 15
Q9XBU9 PRELIMINARY; PRT; 293 AA.
AC Q9XBU9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DB Immunoreactive 32 kDa antigen PG25.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF155351; AAD38410.1; -.
SQ SEQUENCE 293 AA; 32272 MW; CC03EAC241F7F6F1 CRC64;

Query Match
Best Local Similarity 8.1%; Score 213.5; DB 2; Length 293;
Matches 52; Conservative 19; Mismatches 42; Indels 21; Gaps 4;

QY 362 EGVKLPAPYQER-TIDLSAYAGQOVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
Db 39 ESITFVAPVEETAIEAEVLAQEI-----VATBEIA---EQAVRSYTYV 81
QY 421 YRDNVVAQNLAAFTTNOENAVPQYNYCVVEKYTAGVSPKVKDVTVEGSNEFAHYQNL 480
Db 82 YRDGVKIASGLTEPTFLDEDVPAGEHYTCVEVYQGVSDKVCVDVEVK---DFKPTNL 138
QY 481 TGSAVGQKVTWKWD 494
Db 139 TGTASNDVSLDWD 152

Search completed: May 18, 2004, 11:46:34
Job time : 41.3342 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 15.3555 Seconds
(without alignments)
1670.936 Million cell updates/sec

Title: US-08-570-311-2

Perfect score: 2641

Sequence: 1 MRKLSLFLSGLVLLSLLCWG.....QNLGSAVGQKVTLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 2641 | 100.0 | 497 | 2 | US-08-570-311-2 |
| 2 | 2641 | 100.0 | 497 | 2 | US-08-570-311-2 |
| 3 | 2641 | 100.0 | 2628 | 2 | US-08-570-311-14 |
| 4 | 951.5 | 36.0 | 1706 | 4 | US-09-066-330-10 |
| 5 | 944.5 | 35.8 | 1477 | 4 | US-09-482-500A-1 |
| 6 | 944.5 | 35.8 | 1704 | 3 | US-08-336-308A-10 |
| 7 | 944.5 | 35.8 | 1704 | 3 | US-08-822-324-6 |
| 8 | 944.5 | 35.8 | 1704 | 3 | US-09-490-931-10 |
| 9 | 943.5 | 35.7 | 1687 | 2 | US-08-570-311-29 |
| 10 | 816 | 30.9 | 1358 | 2 | US-08-570-311-27 |
| 11 | 812.5 | 30.8 | 1732 | 2 | US-08-570-311-10 |
| 12 | 812.5 | 30.8 | 1732 | 2 | US-08-353-485-10 |
| 13 | 806.5 | 30.5 | 1087 | 2 | US-08-570-311-8 |
| 14 | 806.5 | 30.5 | 1087 | 2 | US-08-353-485-8 |
| 15 | 806.5 | 30.5 | 1732 | 4 | US-09-066-330-11 |
| 16 | 546.5 | 20.7 | 450 | 2 | US-08-570-311-16 |
| 17 | 546.5 | 20.7 | 456 | 2 | US-08-570-311-18 |
| 18 | 546.5 | 20.7 | 456 | 2 | US-08-570-311-20 |
| 19 | 458.5 | 17.4 | 439 | 2 | US-08-570-311-22 |
| 20 | 405 | 15.3 | 942 | 1 | US-08-141-324-14 |
| 21 | 405 | 15.3 | 942 | 1 | US-08-541-902-14 |
| 22 | 175 | 6.6 | 49 | 3 | US-08-822-324-18 |
| 23 | 148 | 5.6 | 42 | 4 | US-09-066-330-3 |
| 24 | 140 | 5.3 | 2736 | 4 | US-09-252-991A-30227 |
| 25 | 135.5 | 5.1 | 2315 | 4 | US-09-543-681A-5434 |
| 26 | 132 | 5.0 | 509 | 3 | US-08-822-324-8 |
| 27 | 127 | 4.8 | 46 | 3 | US-08-822-324-9 |

28 127 4.8 811 1 US-08-480-604A-7 Sequence 7, Appli
29 127 4.8 811 2 US-08-405-496A-7 Sequence 7, Appli
30 127 4.8 811 3 US-08-915-136-7 Sequence 7, Appli
31 127 4.8 811 4 US-08-957-310-7 Sequence 7, Appli
32 127 4.8 811 4 US-10-011-366-7 Sequence 7, Appli
33 127 4.8 811 4 US-09-084-517-7 Sequence 7, Appli
34 127 4.8 812 1 US-08-480-604A-29 Sequence 29, Appli
35 127 4.8 812 3 US-08-915-136-29 Sequence 29, Appli
36 127 4.8 812 4 US-09-084-517-29 Sequence 29, Appli
37 127 4.8 2710 1 US-08-480-604A-6 Sequence 6, Appli
38 127 4.8 2710 2 US-08-405-496A-6 Sequence 6, Appli
39 127 4.8 2710 3 US-08-915-136-6 Sequence 6, Appli
40 127 4.8 2710 4 US-08-957-310-6 Sequence 6, Appli
41 127 4.8 2710 4 US-10-011-366-6 Sequence 6, Appli
42 127 4.8 2710 4 US-09-084-517-6 Sequence 6, Appli
43 126 4.8 1752 4 US-09-865-621A-2 Sequence 2, Appli
44 123 4.7 951 4 US-09-924-097A-15 Sequence 15, Appli
45 121 4.6 25 2 US-08-902-516-46 Sequence 46, Appli

ALIGNMENTS

RESULT 1
US-08-570-311-2
; Sequence 2, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-2

Query Match
Best Local Similarity 100.0%; Score 2641; DB 2; Length 497;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLSLFLSLAVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60
Db 1 MRKLSLFLSLAVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60

QY 61 ARIILEAHWDVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120
Db 61 ARIILEAHWDVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120

QY 121 DASPTNFVLDGTASADIPAGTYDYVIINPNPGIIVVGEVSGKNDYVVEAGKTYHFT 180
Db 121 DASPTNFVLDGTASADIPAGTYDYVIINPNPGIIVVGEVSGKNDYVVEAGKTYHFT 180

QY 181 VORQGPDAASVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVNLSEFDTQ 240
Db 181 VORQGPDAASVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVNLSEFDTQ 240

QY 241 TLPNGWTMIDADGDGHNWLSITNVYNTATHTGCGAMFSKSWTASGGAKIDLSPDNYLVT 300
Db 241 TLPNGWTMIDADGDGHNWLSITNVYNTATHTGCGAMFSKSWTASGGAKIDLSPDNYLVT 300

QY 301 KVTVPENGLSYWVSSQVPTNEHYGVFLSTTGNAAANFTIKLEETLGSDKPAKPNLVK 360
Db 301 KVTVPENGLSYWVSSQVPTNEHYGVFLSTTGNAAANFTIKLEETLGSDKPAKPNLVK 360

QY 361 SEGKLPAPQERTIDLSAVAGQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
Db 361 SEGKLPAPQERTIDLSAVAGQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420

QY 421 YRNVVIAQNLAATTFNQENVAPQYNYCEVKYTAGVSPKVKDVTVEGNEFAHVQNL 480
Db 421 YRNVVIAQNLAATTFNQENVAPQYNYCEVKYTAGVSPKVKDVTVEGNEFAHVQNL 480

RESULT 2
US-08-353-485-2
Sequence 2, Application US/08353485
Patent No. 5830710
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tuwasorn, Somying
APPLICANT: Lepine, Guyline
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-485-2

Query Match
Best Local Similarity 100.0%; Score 2641; DB 2; Length 497;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKLSLFLSLAVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60
Db 1 MRKLSLFLSLAVLLSLCWGQTAAAGGPKTAPSVTHQAVQKGIKRTSKVKDLRDPPIAGM 60

QY 61 ARIILEAHWDVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120
Db 61 ARIILEAHWDVWEDGTGYQMLWDADHNOYGASIPESFWFANGTIPAGLYDPPEYKVPVNA 120

QY 121 DASPTNFVLDGTASADIPAGTYDYVIINPNPGIIVVGEVSGKNDYVVEAGKTYHFT 180
Db 121 DASPTNFVLDGTASADIPAGTYDYVIINPNPGIIVVGEVSGKNDYVVEAGKTYHFT 180

QY 181 VORQGPDAASVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVNLSEFDTQ 240
Db 181 VORQGPDAASVVTGEGNEFAPVQNLQWSVSGQTVTLTWQAPASDKRTYVNLSEFDTQ 240

QY 241 TLPNGWTMIDADGDGHNWLSITNVYNTATHTGCGAMFSKSWTASGGAKIDLSPDNYLVT 300
Db 241 TLPNGWTMIDADGDGHNWLSITNVYNTATHTGCGAMFSKSWTASGGAKIDLSPDNYLVT 300

QY 301 KVTVPENGLSYWVSSQVPTNEHYGVFLSTTGNAAANFTIKLEETLGSDKPAKPNLVK 360
Db 301 KVTVPENGLSYWVSSQVPTNEHYGVFLSTTGNAAANFTIKLEETLGSDKPAKPNLVK 360

QY 361 SEGKLPAPQERTIDLSAVAGQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420
Db 361 SEGKLPAPQERTIDLSAVAGQVYLAFRHFNSTGIFRLYLDVAVSGGSSNDYTYV 420

QY 421 YRNVVIAQNLAATTFNQENVAPQYNYCEVKYTAGVSPKVKDVTVEGNEFAHVQNL 480
Db 421 YRNVVIAQNLAATTFNQENVAPQYNYCEVKYTAGVSPKVKDVTVEGNEFAHVQNL 480

QY 481 TGSVAGQKVTWKWDAPN 497
Db 481 TGSVAGQKVTWKWDAPN 497

RESULT 3
US-08-570-311-14
Sequence 14, Application US/08570311
Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann

APPLICANT: Tumwasorn, Somying
 APPLICANT: Lepine, Guylaine
 APPLICANT: Han, Naiming
 APPLICANT: Lantz, Marilyn
 APPLICANT: Patti, Joseph
 TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
 TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ted W. Whitlock
 STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/570,311
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 PRIOR APPLICATION NUMBER: US 08/353,485
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/647,119
 FILING DATE: 25-JAN-1991
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/241,640
 FILING DATE: 08-SEP-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Whitlock, Ted W.
 REGISTRATION NUMBER: 36,965
 REFERENCE/DOCKET NUMBER: UF15.C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (904) 375-8100
 TELEFAX: (904) 372-5800
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2628 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-570-311-14

Query Match 100.0%; Score 2641; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 1.7e-204;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLNSLFLSLVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKIRTSKVLDLDPACM 60
 DB 1 MKLNSLFLSLVLLSLLCWGQTAAAGGPKTAPSVTHQAVQKIRTSKVLDLDPACM 60
 QY 61 ARILLRAHDVWEDGTGYOMLWDADHNOYXGASIPESFNFANGTIPAGLYDPPEYKVPVNA 120
 DB 61 ARILLRAHDVWEDGTGYOMLWDADHNOYXGASIPESFNFANGTIPAGLYDPPEYKVPVNA 120
 QY 121 DASFSPTNVLGDTASADIPAGTYDYVIINPNPGIYIVGEGVSKGNDVVVEAGKTYHPT 180
 DB 121 DASFSPTNVLGDTASADIPAGTYDYVIINPNPGIYIVGEGVSKGNDVVVEAGKTYHPT 180
 QY 181 VORQPGDAAASVVTGEGNEFAPVONLQWSVSGQVTTLTQAPASDKRTYVLNESFDQ 240
 DB 181 VORQPGDAAASVVTGEGNEFAPVONLQWSVSGQVTTLTQAPASDKRTYVLNESFDQ 240
 QY 241 TLPGNWTMIDADGDGHNWLSLTVNTATHTGDMFMSKSWTASGGAKIDLSPDNLVTP 300
 DB 241 TLPGNWTMIDADGDGHNWLSLTVNTATHTGDMFMSKSWTASGGAKIDLSPDNLVTP 300

QY 301 KVTVPENGKLSYVWSSQVPTNHEHYGVFLSTTCNBAANFTIKLEETLGSDDKDPAPNVLK 360
 DB 301 KVTVPENGKLSYVWSSQVPTNHEHYGVFLSTTCNBAANFTIKLEETLGSDDKDPAPNVLK 360
 QY 361 SEGKVLPAFYQERTIDLSAYAGQOVYLAFRHFNSTGIFRLYLDDVAVSGGSSNDYTYV 420
 DB 361 SEGKVLPAFYQERTIDLSAYAGQOVYLAFRHFNSTGIFRLYLDDVAVSGGSSNDYTYV 420
 QY 421 YRDNVVIQAONLAATTNENQENVAPGOYNYCEVKYTAGVSPKVKCVTVVEGSNEFAHVQNL 480
 DB 421 YRDNVVIQAONLAATTNENQENVAPGOYNYCEVKYTAGVSPKVKCVTVVEGSNEFAHVQNL 480
 QY 481 TGSANGQKVTILKWDAPN 497
 DB 481 TGSANGQKVTILKWDAPN 497
 RESULT 4
 US-09-066-330-10
 ; Sequence 10, Application US/09066330A
 ; Patent No. 6511666
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric C.
 ; APPLICANT: Bhogal, Peter S.
 ; APPLICANT: Stakeski, Nada
 ; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
 ; FILE REFERENCE: Reynolds
 ; CURRENT APPLICATION NUMBER: US/09/066,330A
 ; CURRENT FILING DATE: 1998-09-15
 ; EARLIER APPLICATION NUMBER: PN 6275
 ; EARLIER FILING DATE: 1995-10-30
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00673
 ; EARLIER FILING DATE: 1996-10-30
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 1706
 ; TYPE: PRY
 ; ORGANISM: Porphyromonas gingivalis
 US-09-066-330-10

Query Match 36.0%; Score 951.5; DB 4; Length 1706;
 Best Local Similarity 31.9%; Pred. No. 5.6e-68;
 Matches 237; Conservative 74; Mismatches 156; Indels 275; Gaps 20;
 QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR--TSKVKDLRDPAGMARILAEH 68
 DB 671 TATTGQKVTILKWDAPSTKTATNTARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730
 QY 69 DYWEDGTGYOMLWDADHNOYXGASIPEE--SPWFANGTIPAGLYDPPEYKVPVNAASPSF 126
 DB 731 DYWEDGTGYOMLWDADHNOYXGASIPESDTHTLW--PNCVSPANLFAPEFYTPENADPSCSP 789
 QY 127 TNFVLDGTASADIPAGTYDYVIINPNPGI--IYIVGEGVSKGNDYVVEAGKTYHFTVORQ 185
 DB 790 TNWMDGTASVNIPTASTYDFAIAAPQANAKIWIAGOGTKEDDYVFEAGKTYHFTVORQ 849
 QY 186 PGDAASV-----
 DB 850 SGDGTETLISEGGGSDYTYTVYRDGDKIKEGLTATTFBEDGVATGNEHYCVVEYKVTAGVS 909
 QY 194 -----VTGEGNEFAPVONLQWSVSGQVTTLTQAPAS-----DKRTYVLNE 235
 DB 910 PKVKCVTVVEGSNEFAPVONLQWSVSGQVTTLKWDAPNGTNPENPNPNPGTTTISE 969
 QY 236 SFDTQTLPGNWTMIDADGDGHNWLSLTVNTATHTGDMFMSKSWTASGGAKIDLSPDN 295
 DB 970 SFE--NGIPASWKTIDADGDGHW--KPGNAPGIAGNSNGCVTSESF--GLGGIGV-LTPDN 1025
 QY 236 YLTVPKVTPVNGKLSYVWSSQ--VPTWNEHYGVFLSTTCNBAANFTIKLEETL---GSD 351
 DB 1026 YLITPALDLPNGKLTFWVCQADANYASEHVAVYASSTGNDASFTNALLEETITAKGVR 1085

QY 352 KPAPM----- 356
Db 1086 SPEARGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTTF 1145
QY 357 -----NLVKS----- 361
Db 1146 ESSTHGEAPAEWTTIDADGGQWCLSSGQLDWTAGHGTNVVSPSWNGMALNPDNYL 1205
QY 362 ----- 361
Db 1206 ISKDVGTGATKVKYVAVNDGPGDHVAVMISKTGTNAGDFTVFEETPNGKNGARFGL 1265
QY 362 ----EGVKLPAPYQERTIDLSAYAGQVLYLAFRHFNSGTIFRLYLDV--AVSGEGSND 415
Db 1266 STEADGAKPQSVMIERTVDLP--AGTK-YVAFRHYNSDLYNILLDDIQTWGSPTPTD 1322
QY 416 YTYVYRDNVVIAONLAATTNQENVAPQVNYCVVEKYTAGVSPKCVCKOVTVBGSNEFA 475
Db 1323 YTYVYRDGTIKKEGLTETFEEDGVATGNHEYCVVEKYTAGVSPKCVVTV--NSTQFN 1381
QY 476 HVQNLTGSAVGQKVTWKWDAPN 497
Db 1382 PVKLNKAQPDGDDVVLKWEAPS 1403
RESULT 5
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1
Query Match 35.8%; Score 944.5; DB 4; Length 1477;
Best Local Similarity 31.9%; Pred. No. 1.7e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;
QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR--TSVKQLRDPDPAGMARILLLEAH 68
Db 444 TATTQGGKVLKWDAPSTKTNATNTARSVDGIRELVLLSVSDAPELRGQAEIVLEAH 503
QY 69 DVNEDGTGYQMLDADNQNQASIPEE--SPFWANGTIPAGLYDPFXYKVPVNADASFSP 126
Db 504 DVNNDGSGYQILLDADHDQXQVIPS DTHLW--PNCVSPANLPAPPEYTVPNADPSCSP 562
QY 127 TNFVLDGTASADIPAGYDYVILNPNFGI-IYIVGEGVSKGNDYVVEAGKTYHTVORQG 185
Db 563 TNMTMDGTASVNIIPAGTYVDFALAAPQANAKIWIAGQGPTEDDYVFEAGKKYHFLMKMG 622
QY 186 PGDAASVV----- 193
Db 623 SGDGTETLISEGGSDYTYTVYRDGTIKKEGLTATTFEEDGVATGNHEYCVVEKYTAGVS 682
QY 194 -----VTGEGGNFAPQNLQWSVSGTQVTLTWQAPAS-----DKRTYVLNESF 237
Db 683 PKVKCDTVBGSNEFAFQNLGTSAGVQKVTWKWDAPNGTPNPNPNPNPGITLSESF 742
QY 238 DTQTLPNGWTMIDADGGHWNLSINTVNTATHTGDCAMFSKSWTASGGAKIDLSPDNYL 297

Db 743 E-NCIPASWTKTIDADGGHGW-KPGNAPGAGYNSNGCVYSESF-GUGGTGV-LTPDNYL 798
QY 298 VTPKVTVPENGKLSYVYSSQ-VPMWNEHYGVFLSTTGTGNEAANFTIKLLEETLGS----- 350
Db 799 ITPALDLPNGGKLTFFWCAQDANYASHEVAYASSTGNDASNFTNALLEETITAKGVRSR 858
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 859 EAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTTFES 918
QY 367 ----PAPYQERTID-----LSAYAGQV----- 385
Db 919 STHGEAPAEWTTIDADGGQWCLSSGQLDWTAGHGTNVVAFSFWNGMALNPDNYLIS 978
QY 386 ----- 385
Db 979 KDVGTATKVKYVAVNDGPGDHVAVMISKTGTNAGDFTVFEETPNGKNGARFGLST 1038
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSNDYTYTV 420
Db 1039 EANGAKPQSVMIERTVDLPAGTKYVAFRHYNSDLYNILLDDIQTWGSPTPTDITYTV 1098
QY 421 YRDNVVIAONLAATTNQENVAPQVNYCVVEKYTAGVSPKCVCKOVTVBGSNEFAHYQNL 480
Db 1099 YRDGTIKKEGLTETFEEDGVATGNHEYCVVEKYTAGVSPKCVVTV--NFTQENPVKNL 1157
QY 481 TGSAGVQKVTWKWDAPN 497
Db 1158 KAQPDGDDVVLKWEAPS 1174
RESULT 6
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-308A-10

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
Best Local Similarity 31.9%; Pred. No. 2.1e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSVKDLRDPDPAGMARIIIEAH 68
DB 671 TATTQGGKVTWKWDAPSTKINATNTARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730
QY 69 DWVEDGTGYQMLWDADHNOYGASIPBE--SFWFANGTIPAGLYDPPEYKVPVNDASFSP 126
DB 731 DVWNGSGYQILLDADHDQYGVIPSDTHTLW--PNCVSPANLFAPEYTPVNPADPSCSP 789
QY 127 TNFVLDGTASADIPAGTYDVIINPNPGI--IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185
DB 790 TNMIMOGTASVNPAGTYDFAIAAPOANAKIWIAGQGTKEDDYVFEAGKKIHLMKMG 849
QY 186 PGDAASV-----DKPAPMNLV-----KSEGVKL----- 193
DB 850 SGGTGLTISEGGSDYTYVYRDGTIKIKEGLTATTFEEDGVATGNHEYCVVKYTAGVS 909
QY 194 -----VTGEGNEFAPVQNLQWSVSGQVTTLTWOAPAS-----DKRTYVNLNESF 237
DB 910 PKVKDVTVEGSENEFAPVQNLGTSAGVQKVTWKWDAPNGTPNPNPNPGTTLSESF 969
QY 238 DTQTLPGMTMIDADGHNWLSITINVTATHTGDCGAMESKSWTASGAKIDLSPDNYL 297
DB 970 E-NGIPASWKTIDADGHHGW--KPGNAPGIAGNSNGCVSESF--GLGGIGV--LTPDNYL 1025
QY 298 VTPKVTVPENGKLSYVWSSQ--VPMWNEHYGVFLSTTGNAAANFTIKLEETLGS----- 350
DB 1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRS 1085
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
DB 1086 EAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTFES 1145
QY 367 ----PAPYQERTID-----LSAYAGQV----- 385
DB 1146 STHGAPAEWTTIDADGGQGWLCSSGQLDWTLTHGCTVTVVASFVWNGMALNPDNYLIS 1205
QY 386 ----- 385
DB 1206 KDVGTATKVKYVAVNDGFPGDHYAVVMISKTGTNAGDFTVVFEETPNKNGGARFGLST 1265
QY 386 -----YLAFRHNSGTGIFELYLDVV--AVSGGSSNDYTYV 420
DB 1266 EANGAKPQSWIERTVLDLPAGTKYVAFRHYNCSDLNLYILLDDIQFTWGGSPFTPDYTYV 1325
QY 421 YRDNVVIAQNLAAATTFNQENVAPQYNYCVVEYKTAGVSKVKDVTVEGSENEFAHYNL 480
DB 1326 YRDGTIKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSKVCVNTI--NPTQFPVKNL 1384
QY 481 TGSAGVQKVTWKWDAPN 497
DB 1385 KAOPDGGDVVLKWEAPS 1401

RESULT 7

US-08-822-324-6
Sequence 6, Application US/08822324
Patent No. 6129917
GENERAL INFORMATION:
APPLICANT: Potempa, Jan S.
APPLICANT: Travis, James
APPLICANT: Genco, Caroline A.

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
TITILE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,945
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 103-95 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 488-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-822-324-6

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
Best Local Similarity 31.9%; Pred. No. 2.1e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;

QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSVKDLRDPDPAGMARIIIEAH 68
DB 671 TATTQGGKVTWKWDAPSTKINATNTARSVDGIRELVLLSVSDAPELLRSQAEIVLEAH 730
QY 69 DWVEDGTGYQMLWDADHNOYGASIPBE--SFWFANGTIPAGLYDPPEYKVPVNDASFSP 126
DB 731 DVWNGSGYQILLDADHDQYGVIPSDTHTLW--PNCVSPANLFAPEYTPVNPADPSCSP 789
QY 127 TNFVLDGTASADIPAGTYDVIINPNPGI--IYIVGEGVSKGNDYVVEAGKTYHFTVQROG 185
DB 790 TNMIMOGTASVNPAGTYDFAIAAPOANAKIWIAGQGTKEDDYVFEAGKKIHLMKMG 849
QY 186 PGDAASV-----DKPAPMNLV-----KSEGVKL----- 193
DB 850 SGGTGLTISEGGSDYTYVYRDGTIKIKEGLTATTFEEDGVATGNHEYCVVKYTAGVS 909
QY 194 -----VTGEGNEFAPVQNLQWSVSGQVTTLTWOAPAS-----DKRTYVNLNESF 237
DB 910 PKVKDVTVEGSENEFAPVQNLGTSAGVQKVTWKWDAPNGTPNPNPNPGTTLSESF 969
QY 238 DTQTLPGMTMIDADGHNWLSITINVTATHTGDCGAMESKSWTASGAKIDLSPDNYL 297
DB 970 E-NGIPASWKTIDADGHHGW--KPGNAPGIAGNSNGCVSESF--GLGGIGV--LTPDNYL 1025
QY 298 VTPKVTVPENGKLSYVWSSQ--VPMWNEHYGVFLSTTGNAAANFTIKLEETLGS----- 350
DB 1026 ITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRS 1085
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
DB 1086 EAIRGRIQGTWRQKTVLDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTFES 1145

QY 367 ----PAPYQERTID-----LSAVAGQOV----- 385
Db 1146 STHGEAPAEWTTIDADGGQGWLCSSGQDLTLTAHGGTNVAVSFWNGMALNPDNYLIS 1205
QY 386 ----- 385
Db 1206 KDVGTATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETPNGINKGARFGLST 1265
QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db 1266 EANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOFTWGGSPFTDYTYTV 1325
QY 421 YRDNVVIAQNLAAATFNQENVAPQYNYCVKVTAGVSPKCKDVTVEGSGNEFAHQNL 480
Db 1326 YRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKVTAGVSPKCKCVNVTI-NPTQENPVKNL 1384
QY 481 TGSVAGQKVLKWDAPN 497
Db 1385 KAQPDGGDVVLKWEAPS 1401

RESULT 8

US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US

; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994

; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-490-931-10

Query Match 35.8%; Score 944.5; DB 3; Length 1704;
Best Local Similarity 31.9%; Pred. No. 2.1e-67;

Matches 235; Conservative 75; Mismatches 160; Indels 267; Gaps 20;
QY 22 TABAOG-----GPKTAPSVTHQAVOK--GIR---TSKVKDLRDPPIAGMARIIIEAH 68
Db 671 TATQGGQVTLKWDASTKTNATNTARSVDGRELVLVLSVDAPELRGGQAEIVLEAH 730
QY 69 DVWEDGTGYQMLDADHNOYGASIPBE--SFWFANGTIPAGLYDPPEFYKVPVNADASFSP 126
Db 731 DVWNGSGYQILLDADHDQYGVIPSDTHLW--PNCVSPANLPAPFEYTVENADPSCSP 789
QY 127 TNFVLGTASADIPAGTYDVIINPNPGL--IYIVGEGSVKGNIDYVVEAGTYFTVQROG 185
Db 790 TNMIMDGTASVNPAGTYDFAIAAPOANAKIAGOGPTKEDDYVEAGKYHFLMKMG 849
QY 186 PGDAASVW----- 193
Db 850 SGDGTSLTISEGGGSDYTYVYRDDGKIKKGLTATTFEEDGVATGNHEVCVEVKYTAGVS 909
QY 194 -----VTGEGGNEFAPVQNLWSVSGQTVLLTWQAPAS-----DKRTYVLNESF 237
Db 910 PKVKCDVTVEGSGNEFAPVQNLGSAVGQKVTLAWDAPNGTPNPNPNPNTTILSESF 969
QY 238 DTOTLPGNTMIDADGDGHNWLSITINVTATITGDMFESKSWTASGGAKIDLSPDNYL 297
Db 970 E-NGIPASWKTIDADGDGHW--KPGNAPGIAGYNSGCVYSESF--GLGGTGV-LTPDNYL 1025
QY 298 VTPKVTVPENGKLSYVSSQ--VPWTNEHYGVFLSTTGNAAFTIKLBETLGS----- 350
Db 1026 ITPALDLPNGCKLTFWVCAQADANYASEHYAVYASSTGNDASNTNALLETTITAKVRS 1085
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
Db 1086 EAIRGRIQGTWRQKTVDLDPAGTKYVAFRHFQSTDMFYDLDEVEIKANGKRADETTFES 1145
QY 367 ----PAPYQERTID-----LSAVAGQOV----- 385
Db 1146 STHGEAPAEWTTIDADGGQGWLCSSGQDLTLTAHGGTNVAVSFWNGMALNPDNYLIS 1205
QY 386 ----- 385
Db 1206 KDVGTATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETPNGINKGARFGLST 1265
QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYTYTV 420
Db 1266 EANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOFTWGGSPFTDYTYTV 1325
QY 421 YRDNVVIAQNLAAATFNQENVAPQYNYCVKVTAGVSPKCKDVTVEGSGNEFAHQNL 480
Db 1326 YRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKVTAGVSPKCKCVNVTI-NPTQENPVKNL 1384
QY 481 TGSVAGQKVLKWDAPN 497
Db 1385 KAQPDGGDVVLKWEAPS 1401

RESULT 9

US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville

```
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-29

Query Match 35.7%; Score 943.5; DB 2; Length 1687;
Best Local Similarity 31.8%; Pred. No. 2.5e-67;
Matches 235; Conservative 75; Mismatches 160; Indels 269; Gaps 20;

QY 22 TAAAGG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDIPAGMARILIEAH 68
DB 652 TATTGGKVLKWDAPSTKTATNTARSVDGIRBELVLLSVSDAPELRSQAEIVLEAH 711
QY 69 DVWEDGTGYQMLDADHNDQYASIPER--SFWFANGTIPAGLYDPFFYKVPVNADASFSP 126
DB 712 DVWNDGSGYQILLDADHDQYGVIPSDHTILW-PNCSPANLAFPFYTPVFNADPSCSP 770
QY 127 TNFVLDGTASADIPAGTYDYVINPNPGI-IYIVGEGVSGNDYVVBAGKTYHFTVQROG 185
DB 771 TNMIMDGTASVNPAGTYDYFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKMG 830
QY 186 PGDAASVV----- 193
DB 831 SGDGTETISEGGGSDYTYVYRDGTIKKEGLTATTFEEDGVATGNHEYCVVKYTAGVS 890
QY 194 -----VTGEGGNEFAPVQNLQMSVSGTQVTLTWQAPAS-----DKRTYVLYNE 235
DB 891 PKVKCDVTVGSENEFAPVQNLGTSAGVQKVTLLKWDAPNGTNPENPNPNPNCPTTLSE 950
QY 236 SFTQTLPLNGWTMLDADGDGHNWLTINVTNTATHGTGDGAFKSWTASGAKIDLSPDN 295
DB 951 SFE-NGIPASMKTIIDADGDGHG-KPGNAPGIAGYNGGCVYSESF-GLGGIGV-LTPDN 1006
QY 296 YLVPKTVTPVNGKLSYVWSQ-QVPWTNEHYGFLSTGTGNEAANFTIKLEETLGS----- 350
DB 1007 YLITPDLPLNGGKLTFWCAQDANYASEHYAVYASSTGNDASNFNTALLEETITAKGVR 1066
QY 351 -----DKPAPMNLV-----KSRGVKL----- 366
DB 1067 SPEAIRIGQTRQKIVDLDPAGTKYVAFRHFQSTDMFYDLDELVEIKANGKRADFTTF 1126
```

```
QY 367 -----PAPYQERTID-----LSAYAGQOV----- 385
DB 1127 ESSTHGEAPAEWTTIDADGDGQGLCLSSGQLDMLTAHGGTNVVASFSWNGMALNPDNYL 1186
QY 386 ----- 385
DB 1187 ISKDVGTATKYKYVAVNDGFFGDHYAVMISKTGNTAGDFTVFEETPNGINKGARFGL 1246
QY 386 -----YLAFRHFNSTGIFRLYLDV--AVSGEGSSNDYTY 418
DB 1247 STEANGAKPQSVWIERTVDLPAGTKYVAFRHYVNCSDNLNYILLDDIQTFTMGSPPTDITY 1306
QY 419 TVYRDNVVIAQNLAAATFNQENVAPQYNYCVKYTAGVSPKVKCKDVTVEGSEFAHVQ 478
DB 1307 TVYRDGTYKIKGLTETTFEEDGVATGNHBYCVKYTAGVSPKVCVNTI-NPTQFNPK 1365
QY 479 NLGSAVGQKVTLLKWDAPN 497
DB 1366 NLKAQPDGDDVVLKWEAPS 1384

RESULT 10
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1358 amino acids
```

| TYPE, amino acid | TOPOLOGY: linear | MOLECULE TYPE: protein | US-08-570-311-27 |
|--------------------------------------|---|------------------------|------------------|
| Query Match | 30.8%; Score 816; DB 2; Length 1358; | | |
| Best Local Similarity | 29.6%; Pred. No. 3.7e-57; | | |
| Matches | 219; Conservative 75; Mismatches 171; Indels 274; Gaps 21; | | |
| QY | 22 TAAAGGPKT-----APSVTHQAVQKGIKTSKVKDLRD-----PIPAQMARIILEA 67 | | |
| DB | 328 TATTQGGKVLKWDAPS-----AKKAEASREVKRIKIGDLGVFTVIEPANDVRANEAKVLLAA 382 | | |
| QY | 68 HDVWEDGTGYQMLWDADHNOYASIPESFWFANGTIPAGLYDP--FEYKVPVNADASFSP 126 | | |
| DB | 383 DNVWGDNTGYQFLDADHNTFGSVIPATGPLF--TGTAASNLSANFEYLVPANADPVVTT 441 | | |
| QY | 127 TNFVLDGTASADIPAGTYDYVIINPNP--GIYIVGEG--VSKGNDYVVEAGKTVHFTV 181 | | |
| DB | 442 QNIIVTQGEVWIPGGVYDYCIINPEPASGKMWIAGDGNQPARYDDFTFEAGKKTFTFM 501 | | |
| QY | 182 QROGPGDAASV----- 193 | | |
| DB | 502 RRAAGMGDGTDEVEDDPSASYTYTVYVRDGTIKKEGLTATTFBEDGVAAGNHEYCEVEKVT 561 | | |
| QY | 194 -----VTCEGGNEFAPVQNLQWSVGGQVTLTWOAPAS-----DKRTYVINE 235 | | |
| DB | 562 AGVSPKVKDVTVEGSENEFAPVQNLGSAVGQKVLKWDAPNGTNPNNPFGTTTUSE 621 | | |
| QY | 236 SPTQTLPLNGWTMIDADGDGHNWLSSTINYNTATHTGCGAMFSKWSWTASGGAKIDLSPDN 295 | | |
| DB | 622 SFE--NGIPASWKTIADADGDGHW--KPGNAPGLIAGYNSGCVYSSEF--GLGGIGV-LTPDN 677 | | |
| QY | 296 YLVTPEKVTYPENGKLSWYSSQ--VPWTNHYGVFLSTTCNEAANFTIKLLESTLGS----- 350 | | |
| DB | 678 YLITPALDLANGKLTFWCAQDANTASHEHYAVASSTGNDASNTNALLEETITAKGVR 737 | | |
| QY | 351 -----DKPAPMNLV-----KSEGVKL----- 366 | | |
| DB | 738 SPEAIRGRIQGTWRQKTVLDPAGTKYVAFPHFQSTDMFYDLDLVEIKANGKRAQFTFTF 797 | | |
| QY | 367 -----PAPYQERTID-----LSAYAGQV----- 385 | | |
| DB | 798 ESSTHGEAPAEWTTIDADGDGDQDWLCLSSGOLDWLTAGHGTNTNVASFSWNGMALNPDNYL 857 | | |
| QY | 386 ----- 385 | | |
| DB | 858 ISKDVTKATKVKYXYAVNDGFPDGHVAVMLSKTGNAGDFTVVFETPTENGINKGARGL 917 | | |
| QY | 386 -----YLAFRHNSTGIIFRLYLDV--AVSGEGSSNDYTY 418 | | |
| DB | 918 STEANGAKPQSVWIERTVLDLPAGTKYVAFRHNCSLDLYLLDDIQTFTMGSPPTDVTY 977 | | |
| QY | 419 TVVRDWNVLAAQNLAAFTTQENAVAPQYNNVCEVKYTAGVSKVCKDVTVEGSENEFAHQV 478 | | |
| DB | 978 TVVRDGTIKKEGLTETTFEDGVATGNHHEYCEVKITAGVSKVCNVNVTI--NPTQFNPVK 1036 | | |
| QY | 479 NLGTSVAGQKVLKWDAPN 497 | | |
| DB | 1037 NLKAQPDGDDVVLKWEAPS 1055 | | |
| RESULT 11 | | | |
| US-08-570-311-10 | | | |
| Sequence 10, Application US/08570311 | | | |
| Patent No. 5824791 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Progenle-Fox, Ann | | | |
| APPLICANT: Tumwasorn, Somying | | | |
| APPLICANT: Lepine, Guylaine | | | |
| APPLICANT: Han, Naiming | | | |
| APPLICANT: Lantz, Marielyn | | | |
| APPLICANT: Patti, Joseph | | | |

295 NYLVTPEKVTVPENGKLSYVSSQ--VPWTNEHYGVFLSTGTGNEAANFTIKLEETLGS--- 350
1043 NYLITPALDLPNGGKLTFFWCAQDANYASEHYAVASSIGNDASNTNALLEETITAKGV 1102
351 -----DKPAPMNLV-----KSEGVKL----- 366
1103 RSPKAIRGRIQGTWRQKTVLDPAGTKYVAFRHPQSTDMFYIDLDEVEIKANGKRAFDTET 1162
367 -----PAPYQERTID-----LSAVAGQOV----- 385
1163 FESSTHGEAPAEWTTIDADGGQWLCSSGQDLWLTAGGNSVSSFSWNGMALNPDNY 1222
386 ----- 385
1223 LISKDVTGATKVKYVAVNDGFPDHYAVMISKTGTNAGDFTVPVPEETENGKKGARFG 1282
386 -----YLAFRHNSGTGIFRLYLDV--AVSGEGSNDYT 417
1283 LSTEANGAKPQSVWIBRTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342
418 YTVYRDNVVIAQNLAAATTENQENVAPGQYNYCEVVKYTAGVSPKCKDVTVEGSENEFAHV 477
1343 YTVYRDGTIKKEGLTETFEEDGVAENGHEYCEVVKYTAGVSPKCKVDVTV--NSTQFNPV 1401
478 QNLTSNAV--GQKVTLKWDAP 496
1402 QNLTAQAPNSMDAILKWNAP 1422

RESULT 12

US-08-353-485-10
; Sequence 10, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1732 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-353-485-10

Query Match 30.8%; Score 812.5; DB 2; Length 1732;
Best Local Similarity 30.0%; Pred. No. 1e-56;
Matches 222; Conservative 70; Mismatches 172; Indels 277; Gaps 22;
QY 22 TAAAGGPKT-----APSVTHQVQKIRTSKVKDLRD-----PIPAQMARIILAA 67
DB 693 TATTQKQVTLKWEAPS-----AKAEGSREVKRTGDLGFLVIEPANDVRAEKVVLLA 747
QY 68 HDVWEDGTGYQMLADAHNOYGASIPESFWFANGHIPAGLYDP--FEYKVPVNADASFEP 126
DB 748 DNVWGDNTGYQLLDADHNTFGSVIPATGPLF--TGTAASNLYSANFEYLVANADPVVTT 806
QY 127 TNFVLDDGTASADIPAGTYDYVILNPNP--GIYIIVGEG---VSKGNDYVVEAGKTYHFTV 181
DB 807 QNLIIVTQGEVVIPGGVYDYICITNPEPASGKMWIAGDGGNPARVDDFTFEAGKTYFTM 866
QY 182 ORQGPDAASVV----- 193
DB 867 REAGMGDGTMEVEDDSPAITYTVYRDTGKIKEGLTATTPEEDGVAAGNHEYCEVVKY 926
QY 194 -----VTGEGNEFAPVQNLQMSVSGGTVTLTWOAP-----ASDKRTYVLN 234
DB 927 AGVSPKVKCDVTVEGSENEFAPVQNLTSVSGQKVKLLKWDAPNGTNPNNPNNPGTTL 986
QY 235 ESFDQTLPNGWMTLDADGDGHNLSTNVNTATHTGDGAFSKSWTASGAKIDLSPD 294
DB 987 ESFE--NGIPASWKTIDADGDGHG--KPGNAPGIAGTNSGCVYSBSF--GLGIGV-LTPD 1042
QY 295 NYLVTPEKVTVPENGKLSYVSSQ--VPWTNEHYGVFLSTGTGNEAANFTIKLEETLGS--- 350
DB 1043 NYLITPALDLPNGGKLTFFWCAQDANYASEHYAVVYASSTGNDASNTNALLEETITAKGV 1102
QY 351 -----DKPAPMNLV-----KSEGVKL----- 366
DB 1103 RSPKAIRGRIQGTWRQKTVLDPAGTKYVAFRHPQSTDMFYIDLDEVEIKANGKRAFDTET 1162
QY 367 -----PAPYQERTID-----LSAVAGQOV----- 385
DB 1163 FESSTHGEAPAEWTTIDADGGQWLCSSGQDLWLTAGGNSVSSFSWNGMALNPDNY 1222
QY 386 ----- 385
DB 1223 LISKDVTGATKVKYVAVNDGFPDHYAVMISKTGTNAGDFTVPVPEETENGKKGARFG 1282
QY 386 -----YLAFRHNSGTGIFRLYLDV--AVSGEGSNDYT 417
DB 1283 LSTEANGAKPQSVWIBRTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYT 1342
QY 418 YTVYRDNVVIAQNLAAATTENQENVAPGQYNYCEVVKYTAGVSPKCKDVTVEGSENEFAHV 477
DB 1343 YTVYRDGTIKKEGLTETFEEDGVAENGHEYCEVVKYTAGVSPKCKVDVTV--NSTQFNPV 1401
QY 478 QNLTSNAV--GQKVTLKWDAP 496
DB 1402 QNLTAQAPNSMDAILKWNAP 1422

RESULT 13

US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine

```

; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1087 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-570-311-8
Query Match 30.5%; Score 806.5; DB 2; Length 1087;
Best Local Similarity 29.4%; Pred. No. 1.6e-56;
Matches 211; Conservative 76; Mismatches 166; Indels 265; Gaps 19;

QY 39 AVQKGIKSKVKLRD-----PIPGMARILILEAHVDVWEDGTGYQMLWDADHNOY 88
DB 73 SAKKAERASREVKRGDGLFVTIEPANOVRAEAKVLAADNVGDNVTGYQFLLDADHNTF 132
QY 89 GASIPERSFFANGTIPAGLYDP-FEYKVPVNAADSPFNFLVDGTASADIPAGTYDYV 147
DB 133 GSVIPATGPLE-TGTASNLNYSANFEYLIPANADPVTTQNIIVTGGVEVIFGCVTDYC 191
QY 148 IINPNP-GIITYVGEV---VSKGNDYVVEAGKTYHFTVQRQGGDAASVV-----193
DB 192 ITNPEPAGKMWIAGDGNQAPRYDDFTFEAGKKYFTTMRAGMGDGTDEVEDDSPAY 251
QY 194 -----VTGEGGNEFAP 204
DB 252 TYTVYRDGTKIKGLTATTFEEDGVAAGNHGYCDEVKYTAGVSKVKCKDVTVEGNEFAP 311
QY 205 VQNLQWSVSGVTTLTWOAPAS-----DKRTYVLNESPDTQTLPLNGMTMIDADGDGH 256
DB 312 VQNLTGSAVGKVLKWDAPNGTNPNNPNTGTTLSSEFE-NGIIPASWKTIDADGDGH 370
QY 257 NWLSTINVTATHGTGCMFKSWTASGGAKIDLSPNVLVTPKVTVPENGKLSYVWS 316

```

```

DB 371 GW-KPGNAPGIAGYNSNGCVYSESP-GLGGIGV-LTPDNYLITPALDLANGKGLTFWVCA 427
QY 317 Q-VPTWNEHYGVFLSTGTNEAANFTIKLLEETLGS-----DKP 353
DB 428 QDANYASEHYAVVASSTGNDASGNTFALLEETITAKGVRSPAIRGRIQGTWRQKTVDLDP 487
QY 354 APMLV-----KSEGVKL-----PAPQERTID-----376
DB 488 AGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTTETFESSHGEAPAEWTTIDADGDG 547
QY 377 -----LSAYAGQGV-----385
DB 548 QDWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGF 607
QY 386 -----385
DB 608 PGDHYAVMISKGTGNAGDFTVVFETPNKNGKARFGLSTEANGAKPQSVMIERTVDLP 667
QY 386 ----YLAFFHFNSTGIFRLYLDDV--AVSGESSNDYTYTVYRDNVVIAQNLAAATTENQE 439
DB 668 AGTKYVAFRHYNCSDLDYILLDDIOFTMGGSPTPTDYTYTVYRDGPKIKEGLTETTFEED 727
QY 440 NVAPGOYNYCDEVKYTAGVSPKVKCKDVTVEGSEFAHVNLTGSAVGGKVKYTKWDAPN 497
DB 728 GVATGNHGYCDEVKYTAGVSPKVCVNVTTI-NPTQFNPVKNLKAQPDGDDVVLKWEAPS 784

RESULT 14
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
; APPLICANT: Progluske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1087 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-353-485-8

```

| | | | | |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match | 30.5%; | Score 806.5; | DB 2; | Length 1087; |
| Best Local Similarity | 29.4%; | Pred. No. 1.6e-56; | | |
| Matches 211; | Conservative | 76; | Mismatches 166; | Indels 245; |
| | | | | Gaps 19; |

[illegible]

RESULT 15

US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066.330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275

```

; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

```

| | | | | | | | | |
|-----------------------|-------|--------------|---------|------------|--------|--------|------|----|
| Query Match | 30.5% | Score | 806.5 | DB 4 | Length | 1732 | | |
| Best Local Similarity | 29.7% | Pred. No. | 3.2e-56 | | | | | |
| Matches | 220 | Conservative | 72 | Mismatches | 152 | Indels | 277 | |
| | | | | | | | Gaps | 22 |

| | | | | | | |
|----|------|------------------------|------------------------------|---------------------------|--------------|------|
| QY | 22 | TAAAOQGPKT---- | APSVTHQAVQKGIRTSKVKDLRD----- | PIBAGMARIIIEA | 67 | |
| Db | 693 | TATTQOGKVTLKEAFS---- | AKKAEGSREVKRI | GDGLFVTIEPANDVZANEAKVVLAA | 747 | |
| QY | 68 | HDVWEGTGHOMLWDADHNOY | GASIPESFWFANGTIPAGLYDP- | FEYKVPVNADASFSP | 126 | |
| Db | 748 | DNWVGNTGQFLLDADHNTFGSV | IATGPLF-TGTASSNLYSANFELYLI | PANADPVVTT | 806 | |
| QY | 127 | TNFVLDTASADIPAGTYDVI | INPNP--GIIYIVGEG-- | VSKGNDYVVEAGKTYHFTV | 181 | |
| Db | 807 | QNLIVTGGHEVLPGGVYDICI | NPPEPASKMMIAGDGNQAPARYDDT | FEAGKXYTFTM | 866 | |
| QY | 182 | QROGPGDAASW---- | | | 193 | |
| Db | 867 | RRAGMGDGTMEVEDDSPA | SYTYTVYRDGTKIKEGLTATTT | FEEDGVAAGNHEYCVVEKYT | 926 | |
| QY | 194 | ----- | VTCEGNEFAPQNLQWSVGQV | TTWQAP----- | ASDKRTYVLN | 234 |
| Db | 927 | AGVSPKVKDVTVEGNEFAPQ | NLTGSSVGQVTKLWDAPNGT | PNPNPNPNPGTTLS | 986 | |
| QY | 235 | ESPDQTLPNGTMDADGDGH | NWLSTINVTATHTGDCAMFSK | WTASGGAKIDLSPD | 294 | |
| Db | 987 | ESPE-NGI | PASWKTIADGDGCHW-KPG | NAPGATGNSNGCVVSEF- | GLUGGIV-LTPD | 1042 |
| QY | 295 | NYLIVPKVTPENGKLSYVWSQ- | VEPWTNEHYGVFLSTTGNEA | ANFTIKLEETLGS--- | 350 | |
| Db | 1043 | NYLITPALDLENGGKLTFW | YCAQDANYASEHYAVYASSTG | NDASFTNALLEETITAKGV | 1102 | |
| QY | 351 | ----- | DKPAPMLV----- | KSEGKVL----- | 366 | |
| Db | 1103 | RSPKATRGRIQGTWRQKTVD | LPAGTKYVAFRHQSTDMFYI | DLDEVEIKANGKRADFTET | 1162 | |
| QY | 367 | ----- | PAPYQRTID----- | LSAVAGQOV----- | 385 | |
| Db | 1163 | FESSTHGEAPAEWTTIDADG | GGWMLCSGGQLDWT | TAHGGSNVSVSPWNGMALPN | 1222 | |
| QY | 386 | ----- | | | 385 | |
| Db | 1223 | LISKDVTGATKVXYAVANDG | PPGDHYAVIMISKTGNTAGD | FTVFPETPNGKNGGARFG | 1282 | |
| QY | 386 | ----- | YLAFRHFNSTGIFRLY | ILDDV--AVSGBSSNDYT | 417 | |
| Db | 1283 | LSTEANGAKPQSVMIERTVD | LPAGTKYVAFRHYNCSDLN | YIILLDDIQFTMGGSPTT | 1342 | |
| QY | 418 | YTVYRDNVVIAONLAATTF | ENQENAVPQQNYC | VEVKYTAGVSPKVCXDV | TVESGNEFAHV | 477 |
| Db | 1343 | YTVYRDGTKESGLTETTFE | EDGVATGNHEYC | VEVKYTAGVSPKCKVNTV | -NSTOFPNV | 1401 |
| QY | 478 | QNLTSAV--GOKVTL | KWDAP | 496 | | |
| Db | 1402 | ONLTAEQAPNSMDAILK | WNAP | 1422 | | |

Search completed: May 18, 2004, 11:49:12
Job time : 21.3555 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:42:50 ; Search time 38.497 Seconds
(without alignments)
3592.387 Million cell updates/sec

Title: US-08-570-311-2

Perfect score: 2641

Sequence: 1 MRKINLSFLAVLLSLCWG.....QNLGSAVGQVKTLKWDAPN 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 951.5 | 36.0 | 1706 | 14 | US-10-229-066-10 |
| 2 | 806.5 | 30.5 | 1732 | 14 | US-10-229-066-11 |
| 3 | 791.5 | 30.0 | 419 | 15 | US-10-174-695-3 |
| 4 | 662.5 | 25.1 | 419 | 15 | US-10-174-695-5 |
| 5 | 457.5 | 17.3 | 231 | 15 | US-10-174-695-6 |
| 6 | 186.5 | 7.1 | 196 | 15 | US-10-174-695-4 |
| 7 | 150.5 | 5.7 | 29 | 15 | US-10-387-977-27 |
| 8 | 150.5 | 5.7 | 1483 | 12 | US-10-282-122A-51483 |
| 9 | 148.5 | 5.6 | 42 | 14 | US-10-229-066-3 |
| 10 | 145.5 | 5.5 | 509 | 15 | US-10-387-977-101 |
| 11 | 144.5 | 5.5 | 31 | 15 | US-10-387-977-6 |
| 12 | 143.5 | 5.4 | 1946 | 12 | US-10-282-122A-62947 |
| 13 | 141.5 | 5.3 | 31 | 15 | US-10-387-977-9 |
| 14 | 139.5 | 5.3 | 26 | 15 | US-10-387-977-15 |
| 15 | 138.5 | 5.2 | 2468 | 12 | US-10-282-122A-66335 |

| | | | | | | |
|----|-------|-----|------|----|----------------------|--------------------|
| 16 | 138.5 | 5.2 | 2468 | 14 | US-10-246-330-4 | Sequence 4, Appli |
| 17 | 137 | 5.2 | 25 | 15 | US-10-387-977-82 | Sequence 82, Appl |
| 18 | 137 | 5.2 | 579 | 15 | US-10-369-493-9075 | Sequence 9075, Ap |
| 19 | 134 | 5.1 | 2435 | 12 | US-10-282-122A-47453 | Sequence 47453, A |
| 20 | 132 | 5.0 | 698 | 14 | US-10-156-761-11953 | Sequence 11953, A |
| 21 | 129 | 4.9 | 973 | 14 | US-10-156-761-9394 | Sequence 9394, Ap |
| 22 | 128 | 4.8 | 26 | 15 | US-10-387-977-14 | Sequence 14, Appli |
| 23 | 127 | 4.8 | 811 | 12 | US-10-272-898-7 | Sequence 7, Appli |
| 24 | 127 | 4.8 | 811 | 14 | US-10-011-366-7 | Sequence 7, Appli |
| 25 | 127 | 4.8 | 811 | 15 | US-10-354-774-7 | Sequence 7, Appli |
| 26 | 127 | 4.8 | 811 | 15 | US-10-271-012-7 | Sequence 7, Appli |
| 27 | 127 | 4.8 | 2710 | 12 | US-10-272-898-6 | Sequence 6, Appli |
| 28 | 127 | 4.8 | 2710 | 14 | US-10-011-366-6 | Sequence 6, Appli |
| 29 | 127 | 4.8 | 2710 | 15 | US-10-354-774-6 | Sequence 6, Appli |
| 30 | 127 | 4.8 | 2710 | 15 | US-10-271-012-6 | Sequence 6, Appli |
| 31 | 126 | 4.8 | 1752 | 14 | US-10-387-388-2 | Sequence 2, Appli |
| 32 | 125.5 | 4.8 | 577 | 15 | US-10-369-493-21739 | Sequence 21739, A |
| 33 | 124 | 4.7 | 2013 | 12 | US-10-282-122A-60608 | Sequence 60608, A |
| 34 | 124 | 4.7 | 2358 | 12 | US-10-282-122A-45763 | Sequence 45763, A |
| 35 | 123.5 | 4.7 | 1204 | 12 | US-10-282-122A-49627 | Sequence 49627, A |
| 36 | 123 | 4.7 | 866 | 14 | US-10-222-038-2 | Sequence 2, Appli |
| 37 | 123 | 4.7 | 951 | 9 | US-09-924-097-15 | Sequence 15, Appli |
| 38 | 122 | 4.6 | 806 | 15 | US-10-369-493-3678 | Sequence 3678, Ap |
| 39 | 121 | 4.6 | 25 | 9 | US-09-847-185-46 | Sequence 46, Appli |
| 40 | 121 | 4.6 | 25 | 10 | US-09-930-915A-18 | Sequence 18, Appli |
| 41 | 121 | 4.6 | 25 | 14 | US-10-224-286-46 | Sequence 46, Appli |
| 42 | 121 | 4.6 | 25 | 14 | US-10-062-014-40 | Sequence 40, Appli |
| 43 | 121 | 4.6 | 25 | 14 | US-10-372-076-41 | Sequence 41, Appli |
| 44 | 121 | 4.6 | 25 | 15 | US-10-387-977-76 | Sequence 76, Appli |
| 45 | 121 | 4.6 | 25 | 15 | US-10-387-977-79 | Sequence 79, Appli |

ALIGNMENTS

RESULT 1
US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 36.0%; Score 951.5; DB 14; Length 1706;
Best Local Similarity 31.9%; Pred. No. 3.5e-75;
Matches 237; Conservative 74; Mismatches 156; Indels 275; Gaps 20;
QY 22 TAAAG-----GPKTAPSVTHQAVQK--GIR---TSKVKDLRDPDPAGMALLLEAH 68
DB 671 TATTGGQKVTLLKWDAPSTKTATNTARSVDGIELVLSVSDAPELLRSQGAIVLEAH 730
QY 69 DWEDGTGYQMLWDADHNOYCASIPEE--SPWFANGTIPAGLYDPFFYKVPVNDADSPSP 126
DB 731 DWNDGSGYQILLADHDHQYQGVIPSDTHTLW--PNCSPVNPANLFAFFETVPEADPSCSP 789

QY 127 TNFVLDGTASADIPAGTYDYVIINPNFI-TYIVGEGVSKGNDVYVVEAGKTYHFTVQOR 185
 DB 790 TNMIDGTTASVNI-PAGTYDFAALAPQANAKIWIAGQGPTKEDDYVFEAGKXYHLMKMG 849
 QY 186 PGDAASVV----- 193
 DB 850 SGGDTLTIISGGGSDYTYTVYRDGTIKI EGLTATTFEEDGATGNHBYCVCVKYTAGVS 909
 QY 194 -----VTGEGNEFAPVQNLQWSVSGOTVTLTWQAPAS-----DKRTYVINE 235
 DB 910 PKVKDVTVEGNEFAPVQNLQWSVSGOTVTLTWQAPAS-----DKRTYVINE 235
 QY 236 SFDQTLPNGMTMDADGDGHNWLSSTINVTATHTGDGAMFSGKSWTASGAKIDLSPDN 295
 DB 970 SFE-NGIPASWKTIADGDGHW-KPGNAPAGIAGYNSGCVYSESF-GLGGIGV-LTPDN 1025
 QY 296 YLVTPKVTVPNGKLSYVWSQ-VPMWNEHYGVFLSTTGNENAEFTIKLLETL---GSD 351
 DB 1026 YLIITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLBETITAKGV 1085
 QY 352 KPAPM----- 356
 DB 1086 SPEAMRGRIQGTWRQKTVDLTPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTET 1145
 QY 357 -----NLVKS----- 361
 DB 1146 ESSTHGEAPAEWTITDADGGQWCLSSGOLDMLTAHGGSVNVVSSFSWNGMALNPDNYL 1205
 QY 362 ----- 361
 DB 1206 ISKDVGTATKVKYVAVNDGPGDHYAVMISKTGNAGDFTVVFETPENGINKGARFGL 1265
 QY 362 -----EGVKLPAPYQERTIDLSAYAGQVYLAFRHFNSTGIFRLYLDDV--AVSGEGSSND 415
 DB 1266 STEADGAKPQSWIERTVDLP--AGTK-YVAFRHYNCSDLNILLDDIQFTMGSGSPTPTD 1322
 QY 416 YTYTVYRNVVIAQNLAAATTFNQENVAPQYNYCVCVKYTAGVSPKVKDVTVEGNEFA 475
 DB 1323 YTYTVYRDGTIKI EGLTETTFEEDGATGNHBYCVCVKYTAGVSPKVCVNVTV-NSTQFN 1381
 QY 476 HVONLTGSAVGOKVTLKWDAPN 497
 DB 1382 PVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 2

US-10-229-066-11
 ; Sequence 11, Application US/10229066
 ; Publication No. US20030157637A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric C.
 ; APPLICANT: Bhogal, Peter S.
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
 ; FILE REFERENCE: Reynolds
 ; CURRENT APPLICATION NUMBER: US/10/229,066
 ; CURRENT FILING DATE: 2002-08-28
 ; PRIOR APPLICATION NUMBER: US/09/066,330
 ; PRIOR FILING DATE: 1998-09-15
 ; PRIOR APPLICATION NUMBER: PN 6275
 ; PRIOR FILING DATE: 1995-10-30
 ; PRIOR APPLICATION NUMBER: PCT/AU96/00673
 ; PRIOR FILING DATE: 1996-10-30
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 11
 ; LENGTH: 1732
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-229-066-11

Query Match 30.5%; Score 806.5; DB 14; Length 1732;
 Best Local Similarity 29.7%; Pred. No. 3.2e-62;

Matches 220; Conservative 72; Mismatches 172; Indels 277; Gaps 22;
 QY 22 TAAAGGPKT-----APSVTHQAVQKGIKRTSKVKDLRD-----PIFAGMARILILEA 67
 DB 693 TATTQGGKVTLKWEAPS-----AKKAGSREVKRIKGLFVTTIEPANDVRANEAKVVLAA 747
 QY 68 HDVVEDGTGQMLWDADHNOYGASIPESFPWFANGTIPAGIYDP-FYKVPVVDADAFSP 126
 DB 748 DNVWGDNTGQYQLDADHNTFGSVIPATGPLF-TGTASSNLYSANFEYLLIPANADPVVTT 806
 QY 127 TNFVLDGTASADIPAGTYDYVIINPNP-GIIYIVGEG---VSKGNDVYVVEAGKTYHFTV 181
 DB 807 QNIITVGGEVVPGGVYDYCIITNPEPASGMMIAGDGNQPARYDIDFTFEAGKKTFTM 866
 QY 182 QROGPGDAASVV----- 193
 DB 867 RRACMGDGTMEVEDDSDPASVYTVYRDGTIKI EGLTATTFEEDGVAAGNHEYCVVKYT 926
 QY 194 -----VTGEGNEFAPVQNLQWSVSGOTVTLTWQAP-----ASDKRTYVILN 234
 DB 927 AGVSPKVKDVTVEGNEFAPVQNLQWSVSGOTVTLTWQAP-----ASDKRTYVILN 234
 QY 235 ESDTQTLPLNGMTMDADGDGHNWLSSTINVTATHTGDGAMFSGKSWTASGAKIDLSPD 294
 DB 987 ESFE-NGIPASWKTIADGDGHW-KPGNAPAGIAGYNSGCVYSESF-GLGGIGV-LTPD 1042
 QY 295 NYLVTPKVTVPNGKLSYVWSQ-VPMWNEHYGVFLSTTGNENAEFTIKLLETLGS--- 350
 DB 1043 NYLITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGV 1102
 QY 351 -----DKPAPMNLV-----KSSGVKL----- 366
 DB 1103 RSPKAIRGRIQGTWRQKTVDLTPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRAFDTET 1162
 QY 367 -----PAPYQERTID-----LSAYAGQV----- 385
 DB 1163 PESSTHGEAPAEWTITDADGGQWCLSSGOLDMLTAHGGSVNVVSSFSWNGMALNPDNY 1222
 QY 386 ----- 385
 DB 1223 LISKDVGTATKVKYVAVNDGPGDHYAVMISKTGNAGDFTVVFETPENGINKGARFG 1282
 QY 386 -----YLAFRHFNSTGIFRLYLDDV--AVSGEGSSNDYT 417
 DB 1283 LSTEANGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNILLDDIQFTMGSGSPTPTDT 1342
 QY 418 YTYTVYRNVVIAQNLAAATTFNQENVAPQYNYCVCVKYTAGVSPKVKDVTVEGNEFAH 477
 DB 1343 YTYTVYRDGTIKI EGLTETTFEEDGATGNHBYCVCVKYTAGVSPKVCVNVTV-NSTQFN 1401
 QY 478 QNLTSNAV--GOKVTLKWDAP 496
 DB 1402 QNLTAEQAPNSMDAILKWNAP 1422

RESULT 3

US-10-174-695-3
 ; Sequence 3, Application US/10174695
 ; Publication No. US20030232022A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: Slakeski, Nada
 ; APPLICANT: Chen, Chao Guang
 ; APPLICANT: Barr, Ian George
 ; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
 ; FILE REFERENCE: 52928200700
 ; CURRENT APPLICATION NUMBER: US/10/174,695
 ; CURRENT FILING DATE: 2002-06-18
 ; PRIOR APPLICATION NUMBER: PCT/AU00/01588
 ; PRIOR FILING DATE: 2000-12-21
 ; PRIOR APPLICATION NUMBER: AU PQ 4859
 ; PRIOR FILING DATE: 1999-12-24
 ; NUMBER OF SEQ ID NOS: 8

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match          30.0%; Score 791.5; DB 15; Length 419;
Best Local Similarity 39.9%; Pred.No.8e-62;
Matches 174; Conservative 61; Mismatches 100; Indels 101; Gaps 14;

QY      58 AGMARIIIEAHDVWEDGTGYCMLWDADHNQGASIPEE--SPWFANGTITPAGLYDPEYK 115
Db       :|:||||| |||| |::||| |||| |::||| ::||| |||| |::||| |||| |::|||
            1 SGOAEIVLEAHVDWNDSGGYCILLDAHQDGQQVIPSDTHLW-PNCSSVPANLFAPFEYT 59

QY     116 VPYNADASFSPTNFVLDTGTSADIIPACTDYDVIIINPNPGI-IYIVGEVSXGNNDVVVEAG 174
Db       :|:||||| |||| |::||| |||| |::||| ::||| |||| |::||| |||| |::|||
            60 VPENADPSCSPTNMIMDGTASVNIPAGTYDFALTAAPOANAKIWIAGOGPTYKEDDYVFEG 119

QY     175 KTHFTVQRQGPGDAAHV-----VTCEGNGEFAPNQLQWSVSGQTWTLTWRPAS----- 226
Db       :|:||| :||| :
            120 KCYHFLMKMGSGDGLTELTISEGGSDDYTVVVRDGTKIKEGLTATTFEEDGVGAAGNHXY 179

QY     194 -----VTCEGNGEFAPNQLQWSVSGQTWTLTWRPAS-----                226
Db       :|:||||| |||| |::||| |||| |::||| ::||| |||| |::||| |||| |::|||
            180 CBEVKVTAGVSPKVCKDVTVEGSEFAPQNLTGSAVGQKYLTKWDAPNGTENPNPNP    239

QY     227 --DKRITYLVINESFDOTPLNGWTMIDADGCHNWLSINVNTATHTGDMGFMSKWSTAS 284
Db       :|:||||| |||| |::||| |||| |::||| ::||| |||| |::||| |||| |::|||
            240 NPMPGITLLSEFE-NGIPASWKTDIDAGDHGH-WKNAPGIACIYNSGCCVYESFP-GL   296

QY     285 GKAKILDSPNYLTPKVTYPENGKLISYWSSQ-VPMTNHYHGVLSTTGNEAANTIKL 343
Db       :|:||||| |||| |::||| |||| |::||| ::||| |||| |::||| |||| |::|||
            297 GGIGV-LTPDNLIITPALDLPNGKLTFWCAQDANYASEHYAVVASSTGNDASNFTNAL 355

QY     344 LETILSDEKPAMNLVKSEGKLP-----APYQERTTDLISAYAQOVIAPRHFNST 395
Db       :|:||| :||| :||||| :
            356 LEET-----ITAKVRSPEARIRIQGTWRQKTVDLP--AGTK-YVAFRHFQST 401

QY     396 GIRRLYLDNVAVSSEG 411
Db       :|:||| :||| :
            402 DMFYIDLDEVRIKANG 417


RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match          25.1%; Score 662.5; DB 15; Length 419;
Best Local Similarity 36.0%; Pred.No.2.7e-50;
Matches 157; Conservative 60; Mismatches 118; Indels 101; Gaps 15;
```

QY 254 DGHNLSTINVNTATHTGDMGFMPSKSWTASGGAKIDLSPNYLVTPKVTVPENGKLSYW 313
Db 123 -----
QY 314 VSSQVPWNEHGVFLSTTGNEAANFTIKLEETIGSKDPAPMNLVKSEGVKLPAPYQER 373
Db 123 -----
QY 374 TIDLAVAGQVYLAFRHNSTGIFRLYLDVAVSGEGSSNDYTYTVYRDNVVIAQNLAA 433
Db 128 -----
QY 434 TTFNQENVAPQVNYCVVEKVTAGVSPKVKDVTVEGSENEFAHYQNLGTSAVGQKVLTKW 493
Db 147 TTFEEDGVAAGNHEVCVEKVTAGVSPKVKDVTVEGSENEFAHYQNLGTSSVGQKVLTKW 206
QY 494 DAPN 497
Db 207 DAPN 210

RESULT 6
US-10-174-695-4
; Sequence 4, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match 7.1%; Score 188.5; DB 15; Length 196;
Best Local Similarity 33.5%; Pred. No. 1.8e-08;
Matches 59; Conservative 8; Mismatches 50; Indels 59; Gaps 5;
QY 194 VTGEGNEFAPVQNLQWSVGGQVTLTWQAPAS-----DKRTYVLNSESFTDTQT 241
Db 6 VTGEGNEFAPVQNLQWSVGGQVTLTWQAPNGTTPNPNPNPNPNPNPNPNPNPNPNPNPN 64
QY 242 LKNGWTMDADGDGHNL-----STINVNTATHTGSG 274
Db 65 IPASWTILDGDGCHGKFGNAPGAGYNSGCVYLDNSAKIDRNGEINVNTAEY---- 120
QY 275 AMFSKWTASGGAKIDLSP---DNYLVTPKVTVPENGKLSYVSSQVPTNHYGV 327
Db 121 -----AKTNAPIKVGVYADEKGTGTAAYNWKLSERAKAVAKMKLYGV 164

RESULT 7
US-10-387-977-27
; Sequence 27, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE

; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-27

Query Match 5.7%; Score 150; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 DYTIVYRDNVVIAQNLAAATTFNQENVAP 443
Db 1 DYTIVYRDNVVIAQNLAAATTFNQENVAP 29

RESULT 8
US-10-282-122A-51483
; Sequence 51483, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51483

```
; LENGTH: 1483
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51483

Query Match      5.7%; Score 150; DB 12; Length 1483;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 127; Conservative 81; Mismatches 213; Indels 208; Gaps 29;

QY 27 GGPXTAP-----SYTHQAVQGITSKVKDLRDPDIPAGMARILLLEAHVWEDGTY- 77
Db 522 GTPVVTESQINVEGGSATDQVKLDLNGNTLKQVVDQ-----SGKTLVQGDYTYTDTGIT 577
QY 78 -----QMLWDADHNOYGAS-----IPESEFWFANGTIPAGLYDPPEYKVPV 118
Db 578 LSQSYLAGLALGQYTLILDFNGGASQTIINNVKNETVLSVGTVSGNPGD--TVKVPV 635
QY 119 N-----ADASFPTNF-VLDGTASADIPAGT--YDYVIINPNPGIIVV---- 159
Db 636 TISQVSTPVGILCMDISYDASKFTVKDVLNPTDLVKDNTDYSFVNTSTPGKISITFTDP 695
QY 160 -----GEGVSKNDYVVEAGKT---YHETVQROGPGDAASVVVTGEGNEEPVQNL 208
Db 696 TLANYPIISVGDILAYLDFIINSNATAGDSALT-----DPATLIVADEND-----KDI 743
QY 209 QWSVSGQTVTLTWOAPASDKRTYVLYNESFDT--QTLP-----NGWTMIDA-DGDGH 256
Db 744 KDAASNGKIIVTGSAPV--VQSSVVNTSSVYTDQNAPODQAVSITFNGNTVKDVKDASG- 800
QY 257 NWLSTINVTATHGTGAMFSKWTASGGA-----KIDLSPDN-----YLVTPKVTVPE 306
Db 801 ---NTLKAGSDYATSDGITLSQSYLATLAAGTYTYTIDFSAGNAGTPTVVVKGKTVVGS 857
QY 307 NGKLSYVSS-----QVQPT---NEHYGVFLSTTGNEAANTIK-----LLEET- 347
Db 858 ATTLAVGTVSKAGTVKVPVPTISKVTPVGLICAEIDYDASKFTVKDVLNPTDLVKDTD 917
QY 348 ----LGSDKPAPMNL-----VKSEGVKLPAPYQERTIDLSAYAGQVYLAFRH 391
Db 918 NYSFVNTSTPGKISITFTDLANYPIASDGI---LAYLDFIINSNATAGDSALT--- 971
QY 392 FNSTGIFRLYLDDVAVSGEGSSNDY-----SSVYTDQNAPODQAVSITL 1030
Db 972 NPSGFTIADENDKIDQADANGKITVTGSTPVAENSVVNTSSVYTDQNAPODQAVSITL 1030
QY 418 -----YTVVRDNVIAQNLAAATTENQENVAPQVNYCYEVKY-T 455
Db 1031 NGNTITDVKDASGNTLKAGSDYTVTSDGITLSQSYLAT-----LAAGTYTYTDFDSAGN 1084
QY 456 AGVSPKVKCVTVTEGSENEFAHVQNLTGSA 484
Db 1085 AGTFTVVVKAKTVVSSATTLAVGTVSGKA 1113

RESULT 9
US-10-229-066-3
; Sequence 3, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-3

Query Match      5.6%; Score 148; DB 14; Length 42;
Best Local Similarity 64.1%; Pred. No. 7.3e-06;
Matches 25; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 58 AGMARILLEADHVDWEGTGYQMLWDADHNOYGASIPES 96
Db 1 SQGAIEVLEADHVDWNGSGYQILLDADHDQYGVIPSDT 39

RESULT 10
US-10-387-977-101
; Sequence 101, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-101

Query Match      5.5%; Score 145; DB 15; Length 509;
Best Local Similarity 19.1%; Pred. No. 0.00061;
Matches 102; Conservative 65; Mismatches 155; Indels 212; Gaps 22;

QY 114 YKVPVN---ADASF-----SPTNFVLD-----GTASADIPAGTY----D 145
Db 10 YNTPVRLVWAGAKFKALKPWLTKAQKGFYLDVHYTDEAEVGTTNASIKAFTHKKYND 69
QY 146 YVIINPNPGIYIVGEGVSKNDYVVEAGKTYHETVQROGPGDAASVVVTGEGNEFAPV 205
Db 70 GLAASAAPVFLALVGD-----TDVISGEKSKTKYV 100
QY 206 QNLQWS-VSG---QTVTLTWOAPASDKRTYVLYNE--SFDTQTLPN---GWTMIDADGD 254
Db 101 TDLYSAVDGDYFFPEMYTFRMSASSPEELTWIIDKVLMEYKATPDKSYLKVLLIAGAD 160
QY 255 GHNWLSTI-----NYVN-----TATHYGDGAMFSKWTASGG 286
Db 161 -YSNWSQVGOPTIKYGMQYVYNNQEHGYTDVNYLKPATYTCYSHLNTGVSPA-NYTAGS 218
QY 287 A-----KIDLSPDNYLVTPK-----VTVPEKGLSYWV 314
Db 219 ETAWADPLLTTSQKALTNKDKYELAIQNCCTIAQFYDVQPCFGEVITRVEKEGAYAYIG 278
QY 315 SSQVPTWNEHY-----GVFLSTTGNEAANFTIKLLEET----- 347
Db 279 SSPNSYWGEDIYWSVGANAVGVQVPTFEGTSMGSDATFLEDSTYNTVNSIMWAGNLAATH 338
QY 348 -----LGSDKPAPMN-LVKSEGVKLPV---PYQERTIDLSAYAGQ 383
```

Db 339 AGNIGNTHICAHYWEAYHVLGDSVMPRAEMFKTYTILPASLPQNASYSIQASGS 398
Qy 384 QVYLAFRHNSTGIFRLYLDVAVSGSSNDYTYTYVYRDNVVIQAOLAAATFQENVAP 443
Db 399 Y-----VAISKD-----VLYGTGVANASGVATVSMTKQITEN 431
Qy 444 GOYNYCUEVKYTAGVSPKVKCDVTEGSENEFAHVQNLTGSAVGOKVTLKWDAPN 497
Db 432 GNY-----DWITRSNYLPVKIQVGBESPYPQVSNLTATTQOKVKLKEAPS 481

RESULT 11

US-10-387-977-6
; Sequence 6, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-6

Query Match 5.5%; Score 144; DB 15; Length 31;
Best Local Similarity 96.6%; Pred. No. 1e-05;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 469 EGSNEFAHVQNLTGSAVGOKVTLKWDAPN 497
Db 1 EGSNEFAPVQNLTGSAVGOKVTLKWDAPN 29

RESULT 12

US-10-282-122A-62947
; Sequence 62947, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62947
; LENGTH: 1946
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-62947

Query Match 5.4%; Score 143; DB 12; Length 1946;

Best Local Similarity 18.4%; Pred. No. 0.0071;
Matches 110; Conservative 63; Mismatches 214; Indels 210; Gaps 23;

Qy 22 TAAAGGKFTAPSVTHQAVQKIRTSKVKDLRD-----PIAGMARIILEAHDVWEDGTGY 77
Db 417 TATLGGGLTVKDIHNKQIQVSADGKIKFTDNGGTSAPVAAGTTRIT-----RDTIGF 469
Qy 78 QMLWDADHNQYCASIPESFPFANGTIPAGIYDFEYKVPVNADASPTNFVLDDG-TAS 136
Db 470 A-----KEGTVDDNSPYLDNAKLKVG-----NVEINKDTGGINAGKQITGLTGD 514
Qy 137 ADIPAGTYDYVIINPNPGIYIVGEGVSKGNDYVVEAGKYHFTVQROGPGDAASVVVTG 196
Db 515 ADDADAVTIKQAKAKNL--TPGNGIEINN-----TNSLLVDA 551
Qy 137 EGGNEFAPVQNLQMSVSGQTVTLTWOAPA-----SDKRTY-----VINESFD 238
Db 552 ANGNVTPP-----SYTIGVKTTTLTNNGPSDSKFVAKDSNTHNNSLVTAADLAGYLNEYNR 607
Qy 239 T-QTLPGWTMIDADGDGHNLSTINVTATHTGDMGAFSKSWTASGAKIDLSPDNYL 297
Db 608 TADSALQSFVTKGGDAASNSITLIDKTSNSLTVAGDNGITVKTDTTKKVTVGIDQANGL 667
Qy 298 VTPKVTVPENGKLSYVWSSQVPTNEHYGVFLSTTGNAAFTIKLLEETL----- 348
Db 668 TPKLTVGNSKTLQVIEQMV-----IGNDTKNI-IKGLSSTLTIDTNTDNT 713
Qy 349 -----GSDKPA-----PMNLVKSQGVKLPAPYQ 371
Db 714 HTTEQDNDAAQKESNAASIKDVLNAGFLQNGKPFVDFVSTYDVTNPFANGDGTATVTV- 772
Qy 372 ERTIDLSAVAGQVYLAFRHEN-----STGIFRLYLDVAV-----SGEGSSNDYTYTV 421
Db 773 -----NYNEGKPTSTVTVVNVVDDTTIHLANNNGNKLSVKTTILT 813
Qy 422 RDNVVIQNLAAATTF-----NOENVAPGQVNYCUEVKYTAGVSP-----KVCKD- 465
Db 814 KIN---GANGNATKFSANNGDALVNAKGIADNLNLILABEIHNTTKGTADTALQTFKVKDGG 870
Qy 466 -----VTVEGSNEFA-----HVQNLTCGSAVGOKVTLKWD 494
Db 871 ATNDDDDTTITVGKNAVDTLTFKGENGLTVATKDGDTVTFTGINTQSGLKAGDNTLNKD 927

RESULT 13

US-10-387-977-9

```
; Sequence 9, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-10-387-977-9

Query Match          5.3%; Score 141; DB 15; Length 31;
Best Local Similarity 93.1%; Pred. No. 1.9e-05;
Matches 27; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 469 EGSNEFAHVQNLTSAGVQKVKTLKWDAPN 497
Db 1 EGSNEFAPVQNLTSAGVQKVKTLKWDAPN 29

RESULT 14
US-10-387-977-15
; Sequence 15, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-10-387-977-15

Query Match          5.3%; Score 139; DB 15; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 FAPVQNLQWSVSGQTIVLTWQAPASD 227
Db 1 FAPVQNLQWSVSGQTIVLTWQAPASD 26
```

RESULT 15

```
US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66335
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-10-282-122A-66335

Query Match          5.2%; Score 138.5; DB 12; Length 2468;
Best Local Similarity 23.7%; Pred. No. 0.026;
Matches 107; Conservative 47; Mismatches 166; Indels 131; Gaps 23;

Qy 103 TIPAGLYDPFVKVPVNADASFPTNEVLDTGTSADIPAGTYDYVILNPNPGIYIVGEG 162
Db 319 TDPAG-----NNSTPVTVEAP-----DTTA-----PAPATD-----VQVAPDG 351
Qy 163 VS-KGNDYVVEAGKTYHFTVQROGPGDAASVVVTGEGNEFAPVQNLQWSVSGQTIVLTW 221
Db 352 SSVTGN---AEPGAT---VGVDTDGQDPDPTVVVVGPGSGFEVPLN---PPIITNGETVTVIV 404
Qy 222 QAPASDKRTYVLMNESFDITQLPN-----GWTMIDADGDGH----- 256
Db 405 TDPAGNSSTPTVTAEPDPDPAPQVNASNGSVLSGTAEAGVTIVITDGNPNIGQTSADAN 464
Qy 257 -NMLST-----INVY-----NTATH-TGDGAM-----FKSWTASG 285
Db 465 GNNSTFGSQLPDGTVVNVVARDAGNSSPATISITVDGVAPNAPVVEPFSNGSELSTGTAEP 524
Qy 286 GAKIDLSPDNLYLTPKVTVPENGKLSYVSSQVPTNEHYGVFLSTTGTGNEAANFTIKLLE 345
Db 525 GSSVTLTLDGNGNPIGQTTADANGNWSFTPTPLP-----DGTVVNVVARDAGNSSPAS 579
```

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-----|------------|-------|------|-----|------|------|-------|-------|-----|------|-------|------|----|----|-----|-----|----|----|--------|-----|----|---|-----|----|-----|----|---|---|---|-----|---|---|---|---|---|-----|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| Qy | 346 | ETLGSDDKPA | PNLKV | SEGV | KLP | APYQ | ERTI | DL | SAY | AGQ | VYLA | PRHFN | STGI | FR | LY | LDV | 405 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 580 | VTVD | AV | AT | PT | VD | PS | NGT | ----- | TL | SG | TA | EP | GS | SV | TL | TD | GN | PI | G----- | 621 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 406 | AVS | EG | SS | ND | Y | T | ----- | VY | RD | VV | IA | Q | N | LA | AT | T | PN | Q | EN | V | AP | G | Q | NY | C | VE | K | Y | T | A | G | V | S | P | K | 461 | | | | | | | | | | | | | | | | | | | |
| Db | 622 | Q | V | T | A | D | G | S | G | N | - | W | T | F | T | P | S | T | P | L | P | N | G | T | V | --- | N | A | T | A | T | D | P | S | G | N | A | S | S | F | A | S | ----- | V | T | V | D | A | V | A | P | A | T | P | V | 673 |
| Qy | 462 | V | - | - | C | K | D | V | T | V | E | G | S | N | E | F | A | H | V | O | N | I | T | --- | G | S | A | V | G | Q | 487 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 674 | V | N | P | S | N | G | T | I | L | S | G | T | A | E | P | C | A | T | V | T | L | D | G | N | E | P | I | G | Q | 704 | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: May 18, 2004, 11:52:19
Job time : 40.497 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:33:39 ; Search time 46.4099 Seconds

(without alignments)
2739.638 Million cell updates/sec

Title: US-08-570-311-16

Perfect score: 2443

Sequence: 1 PNPNGTTLTSEFENGIPA.....QNLTGSAGVGKVLKWDAPN 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|--------|---------|--------------|------------|--------------------|
| 1 | 2443 | 100.0 | 450 | 2 AAR96021 | Aar96021 P. gingiv |
| 2 | 2443 | 100.0 | 450 | 2 AAW69489 | Aaw69489 Haemagglu |
| 3 | 2443 | 100.0 | 450 | 2 AAR96030 | Aar96030 P. gingiv |
| 4 | 2443 | 100.0 | 2628 | 2 AAW69488 | Aaw69488 Haemagglu |
| 5 | 2436 | 99.7 | 456 | 2 AAR96023 | Aar96023 P. gingiv |
| 6 | 2436 | 99.7 | 456 | 2 AAR96022 | Aar96022 P. gingiv |
| 7 | 2436 | 99.7 | 456 | 2 AAW69491 | Aaw69491 Haemagglu |
| 8 | 2436 | 99.7 | 456 | 2 AAW69490 | Aaw69490 Haemagglu |
| 9 | 2321 | 95.0 | 439 | 2 AAR96024 | Aar96024 P. gingiv |
| 10 | 2321 | 95.0 | 439 | 2 AAW69492 | Aaw69492 Haemagglu |
| 11 | 2063 | 84.4 | 1687 | 2 AAR96033 | Aar96033 P. gingiv |
| 12 | 2063 | 84.4 | 1687 | 2 AAW69495 | Aaw69495 Haemagglu |
| 13 | 2063 | 84.4 | 1704 | 2 AAR70188 | Aar70188 Arg-gingi |
| 14 | 2063 | 84.4 | 1704 | 2 AAW34843 | Aaw34843 Arg-gingi |
| 15 | 2063 | 84.4 | 1704 | 3 AAR67396 | Aar67396 Arg-gingi |
| 16 | 2063 | 84.4 | 1704 | 4 AAU08938 | Aau08938 P. gingiv |
| 17 | 2051 | 84.0 | 1706 | 2 AAW24786 | Aaw24786 PrtR anti |
| 18 | 2036 | 83.3 | 1087 | 2 AAR96028 | Aar96028 P. gingiv |
| 19 | 2036 | 83.3 | 1087 | 2 AAW69486 | Aaw69486 Haemagglu |
| 20 | 2036 | 83.3 | 1358 | 2 AAR96032 | Aar96032 P. gingiv |
| 21 | 2036 | 83.3 | 1358 | 2 AAW69494 | Aaw69494 Haemagglu |
| 22 | 2011.5 | 82.3 | 1732 | 2 AAR96029 | Aar96029 P. gingiv |
| 23 | 2011.5 | 82.3 | 1732 | 2 AAW24787 | Aaw24787 PrtK anti |
| 24 | 2011.5 | 82.3 | 1732 | 2 AAW69487 | Aaw69487 Haemagglu |
| 25 | 727 | 29.8 | 135 | 6 ABP55081 | Abp55081 Porphyrom |

ALIGNMENTS

RESULT 1

AAR96021
ID AAR96021 standard; protein; 450 AA.

XX AC AAR96021;

XX XX

DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX XX

DE P. gingivalis haemagglutinin hAgA Harepl product.

XX KW

Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harepl.

XX OS Porphyromonas gingivalis; strain 381.

XX FN WO9617936-A2.

XX XX

PD 13-JUN-1996.

PP 11-DEC-1995; 95WO-US016108.

XX XX

PR 09-DEC-1994; 94US-00353485.

XX XX

PA (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX XX

PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX XX

DR WPI; 1996-287181/29.

XX N-PSDB; AAT30645.

XX PT Porphyromonas gingivalis genes and proteins - used in the detection and

XX vaccination against periodontal disease.

XX XX

PS Claim 4; Page 103-104; 153pp; English.

XX XX

CC Harepl (AAR96021) is the product of the Harepl repeat unit (AAT30645) of

CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of

CC haemagglutinin hAgA (see also AAR96030). Harepl and other hAgA repeat

CC unit products (see also AAR96022-24) can be obtd. from transformed host

CC cells and used as vaccines to protect humans or animals against

CC periodontal disease. Expression in Salmonella cells allows prodn. of live

CC vaccine. Harepl-4 can also be used to detect the presence of anti-P.

CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic

XX appn. (Updated on 16-Oct-2003 to standardise OS field)

XX SQ Sequence 450 AA;

Aab49217 Peptide u
Aau03572 P. gingiv
Aay34522 Porphyrom
Aay34521 Porphyrom
Aay34520 Porphyrom
Aay34392 Porphyrom
Aau03574 P. gingiv
Aau03574 P. gingiv
Aay34359 Porphyrom
Aay34484 Porphyrom
Aar96025 P. gingiv
Aaw69483 Haemagglu
Aar72458 Porphyrom
Aau03575 P. gingiv
Aar77313 Porphyrom
Aaw34805 Arg-speci
Aay34483 Porphyrom
Aay34358 Porphyrom
Aau03573 P. gingiv
Aaw34798 Arg-speci
Aaw83085 Peptide f

26 711 29.1 134 4 AAB49217
27 700 28.7 419 4 AAU03572
28 683 28.0 921 2 AAY34522
29 683 28.0 922 2 AAY34521
30 683 28.0 925 2 AAY34520
31 683 28.0 938 2 AAY34392
32 680.5 27.9 419 4 AAU03574
33 565 23.1 377 2 AAY34359
34 561.5 23.0 312 2 AAY34484
35 546.5 22.4 497 2 AAR96025
36 546.5 22.4 497 2 AAW69483
37 415 17.0 970 4 AAR72458
38 385 15.8 231 4 AAU03575
39 267 10.9 991 2 AAR77313
40 236 9.7 49 2 AAW34805
41 233 9.5 293 2 AAY34483
42 233 9.5 299 2 AAY34358
43 206.5 8.5 196 4 AAU03573
44 189 7.7 46 2 AAW34798
45 153 6.3 29 2 AAW83085

Query Match 100.0%; Score 2443; DB 2; Length 450;
 Best Local Similarity 100.0%; Pred. No. 1.2e-184;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNPNGPTTLLSESPENGIPASWKTTIDADGDNNTTTPPGGTSPAGHNSAICASSAYI 60
 DB 1 PNPNGPTTLLSESPENGIPASWKTTIDADGDNNTTTPPGGTSPAGHNSAICASSAYI 60

QY 61 NFEQPNDNVLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 DB 61 NFEQPNDNVLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

QY 121 BEVLTAKTVTTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
 DB 121 BEVLTAKTVTTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

QY 181 ANGKRADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQDLWLTAGGNTNVVASFS 240
 DB 181 ANGKRADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQDLWLTAGGNTNVVASFS 240

QY 241 WNGMALNPNDNVLISKDVTGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 300
 DB 241 WNGMALNPNDNVLISKDVTGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 300

QY 301 NGINKGARFGLSTADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNILLDDIOFT 360
 DB 301 NGINKGARFGLSTADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNILLDDIOFT 360

QY 361 MGSPTPTDYYTVVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECNV 420
 DB 361 MGSPTPTDYYTVVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECNV 420

QY 421 TVDPVQFNPVQNLTSAGVQKVTLLKWDAPN 450
 DB 421 TVDPVQFNPVQNLTSAGVQKVTLLKWDAPN 450

RESULT 2

AAW69489
 ID AAW69489 standard; protein; 450 AA.
 XX AC AAW69489;
 XX DT 22-DEC-1998 (first entry)
 XX DE Haemagglutinin protein hgaA, Harepl.
 XX KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX OS Porphyromonas gingivalis.
 XX PN US5824791-A.
 XX PD 20-OCT-1998.
 XX PF 11-DEC-1995; 95US-00570311.
 XX PR 08-SEP-1988; 88US-00241640.
 XX PR 25-JAN-1991; 91US-00647119.
 XX PR 09-DEC-1994; 94US-00353485.
 XX XX (UYFL) UNIV FLORIDA.
 XX PA (UABR-) UAB RES FOUND.
 XX PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
 XX PI WPI; 1998-582627/49.
 XX DR N-PSDB; AAV58876.
 XX XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 XX PT protease poly.peptide(s)).
 XX XX Claim 1; Col 121-126; 101pp; English.

XX This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease

SQ Sequence 450 AA;

Query Match 100.0%; Score 2443; DB 2; Length 450;
 Best Local Similarity 100.0%; Pred. No. 1.2e-184;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNPNGPTTLLSESPENGIPASWKTTIDADGDNNTTTPPGGTSPAGHNSAICASSAYI 60
 DB 1 PNPNGPTTLLSESPENGIPASWKTTIDADGDNNTTTPPGGTSPAGHNSAICASSAYI 60

QY 61 NFEQPNDNVLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 DB 61 NFEQPNDNVLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120

QY 121 BEVLTAKTVTTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
 DB 121 BEVLTAKTVTTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180

QY 181 ANGKRADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQDLWLTAGGNTNVVASFS 240
 DB 181 ANGKRADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQDLWLTAGGNTNVVASFS 240

QY 241 WNGMALNPNDNVLISKDVTGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 300
 DB 241 WNGMALNPNDNVLISKDVTGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 300

QY 301 NGINKGARFGLSTADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNILLDDIOFT 360
 DB 301 NGINKGARFGLSTADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNILLDDIOFT 360

QY 361 MGSPTPTDYYTVVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECNV 420
 DB 361 MGSPTPTDYYTVVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECNV 420

QY 421 TVDPVQFNPVQNLTSAGVQKVTLLKWDAPN 450
 DB 421 TVDPVQFNPVQNLTSAGVQKVTLLKWDAPN 450

RESULT 3

AAW6030
 ID AAR96030 standard; protein; 2628 AA.
 XX AC AAR96030;
 XX DT 16-OCT-2003 (revised)
 XX DT 04-SEP-1996 (first entry)
 XX DE P. gingivalis haemagglutinin hgaA.
 XX KW Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.
 XX OS Porphyromonas gingivalis; strain 381.
 XX FH Key Location/Qualifiers
 XX FT Peptide 5..21
 XX FT /label= Sig_peptide
 XX XX WO9617936-A2.
 XX PN 13-JUN-1996.
 XX PD 11-DEC-1995; 95WO-US016108.
 XX PF 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Progulsk-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX WPI; 1996-287181/29.
 DR N-PSDB; AAT30654.
 XX
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 6; Page 93-101; 153pp; English.
 XX
 CC P. gingivalis 381 haemagglutinin hgaA (AAR36030) was identified as the
 CC product of the hgaA gene (AAR30654) isolated as an EcoRV fragment of
 CC genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
 CC 24), can be obtd. from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
 CC also be used to detect the presence of anti-P. gingivalis antibodies and
 CC to raise monoclonal antibodies for diagnostic appin. (Updated on 16-OCT-
 CC 2003 to standardise OS field)
 XX
 SQ Sequence 2628 AA;

Query Match 100.0%; Score 2443; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 1.4e-183;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PNPNGTTTLLSESPENGIPASWKTIDADGCGNNWTTTTPPGGTSFAGHNSAICASSAYI 60
 DB 500 PNPNGTTTLLSESPENGIPASWKTIDADGCGNNWTTTTPPGGTSFAGHNSAICASSAYI 559
 QY 61 NFEQPQNDNYLVTPPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 DB 560 NFEQPQNDNYLVTPPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619
 QY 121 EEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINLDDVEIK 180
 DB 620 EEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINLDDVEIK 679
 QY 181 ANGKRAPFTTFESSTHGEAPAEWTTTIDADGCGGWLCLSSGQLDWTLAGHGTNVVASFS 240
 DB 680 ANGKRAPFTTFESSTHGEAPAEWTTTIDADGCGGWLCLSSGQLDWTLAGHGTNVVASFS 739
 QY 241 WNGMALPNPNYLI SKDVTGATKVKYVAVNDGFFGDHYAVMISKTGTNAGDFTVVFBEPT 300
 DB 740 WNGMALPNPNYLI SKDVTGATKVKYVAVNDGFFGDHYAVMISKTGTNAGDFTVVFBEPT 799
 QY 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOFT 360
 DB 800 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOFT 859
 QY 361 MGSPTPTDITYTVYRGTGKIKEGLTTFEEDGVAIGNHGYCVVEKYTAGVSPKCVNV 420
 DB 860 MGSPTPTDITYTVYRGTGKIKEGLTTFEEDGVAIGNHGYCVVEKYTAGVSPKCVNV 919
 QY 421 TVDPVQNFVQNLTGSAVGQKVTLLKWDAPN 450
 DB 920 TVDPVQNFVQNLTGSAVGQKVTLLKWDAPN 949

RESULT 4
 AAW69488
 ID AAW69488 standard; protein; 2628 AA.

XX AAW69488;
 XX
 XX 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hgaA.

KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX Porphyromonas gingivalis.
 OS
 XX US5824791-A.
 XX 20-OCT-1998.
 XX
 XX 11-DEC-1995; 95US-00570311.
 XX
 XX 08-SEP-1988; 88US-00241640.
 XX 25-JAN-1991; 91US-00647119.
 XX 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 XX (UABR-) UAB RES FOUND.
 XX
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulsk-Fox A, Lepine G;
 XX WPI; 1998-582627/49.
 XX N-PSDB; AAV58875.
 XX
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 XX protease poly:peptide(s)).
 XX
 XX Claim 1; Col 91-110; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 SQ Sequence 2628 AA;

Query Match 100.0%; Score 2443; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 1.4e-183;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PNPNGTTTLLSESPENGIPASWKTIDADGCGNNWTTTTPPGGTSFAGHNSAICASSAYI 60
 DB 500 PNPNGTTTLLSESPENGIPASWKTIDADGCGNNWTTTTPPGGTSFAGHNSAICASSAYI 559
 QY 61 NFEQPQNDNYLVTPPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 DB 560 NFEQPQNDNYLVTPPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 619
 QY 121 EEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINLDDVEIK 180
 DB 620 EEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINLDDVEIK 679
 QY 181 ANGKRAPFTTFESSTHGEAPAEWTTTIDADGCGGWLCLSSGQLDWTLAGHGTNVVASFS 240
 DB 680 ANGKRAPFTTFESSTHGEAPAEWTTTIDADGCGGWLCLSSGQLDWTLAGHGTNVVASFS 739
 QY 241 WNGMALPNPNYLI SKDVTGATKVKYVAVNDGFFGDHYAVMISKTGTNAGDFTVVFBEPT 300
 DB 740 WNGMALPNPNYLI SKDVTGATKVKYVAVNDGFFGDHYAVMISKTGTNAGDFTVVFBEPT 799
 QY 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOFT 360
 DB 800 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOFT 859
 QY 361 MGSPTPTDITYTVYRGTGKIKEGLTTFEEDGVAIGNHGYCVVEKYTAGVSPKCVNV 420
 DB 860 MGSPTPTDITYTVYRGTGKIKEGLTTFEEDGVAIGNHGYCVVEKYTAGVSPKCVNV 919
 QY 421 TVDPVQNFVQNLTGSAVGQKVTLLKWDAPN 450
 DB 920 TVDPVQNFVQNLTGSAVGQKVTLLKWDAPN 949

```

RESULT 5
AAR96023
ID AAR96023 standard; protein; 456 AA.
XX
XX AAR96023;
XX
XX 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
XX P. gingivalis haemagglutinin hgaA Harep3 product.
DE
XX
XX Haemagglutinin; hgaA; periodontal disease; vaccine; antibody; Harep3.
XX
XX Porphyromonas gingivalis; strain 381.
XX
XX WO9617936-A2.
XX
XX 13-JUN-1996.
XX
XX 11-DEC-1995; 95WO-US016108.
XX
XX 09-DEC-1994; 94US-00353485.
XX
XX (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
XX Progulske-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
XX WPI; 1996-287181/29.
DR N-PSDB; AAT30647.
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
XX
XX Claim 4; Page 110-112; 153pp; English.
XX
XX Harep3 (AAR96023) is the product of the Harep3 repeat unit (AAT30647) of
CC the hgaA gene (AAT30654) of P. gingivalis 318. It forms part of
CC haemagglutinin hgaA (see also AAR96030). Harep3 and other hgaA repeat
CC unit products (see also AAR96021-22 and AAR96024) can be obtd. from
CC transformed host cells and used as vaccines to protect humans or animals
CC against periodontal disease. Expression in Salmonella cells allows produ.
CC of live vaccine. Harep1-4 can also be used to detect the presence of anti
CC -P. gingivalis antibodies and to raise monoclonal antibodies for
CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 456 AA;

Query Match 99.7%; Score 2436; DB 2; Length 456;
Best Local Similarity 99.8%; Pred. No. 4.2e-184;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PNPNGPTTILSFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICASSAYI 60
DB 7 PNPNGPTTILSFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICASSAYI 66
QY 61 NFEQPQNPNDYLTPELSIPNGGTLTFWVCAQDANYASHYAVYASSTGNDASNANALL 120
DB 67 NFEQPQNPNDYLTPELSIPNGGTLTFWVCAQDANYASHYAVYASSTGNDASNANALL 126
QY 121 BEVLTAKTVVTAPEAIRGTRVQGTQKTVQLPAGTKYVAFRHFCTDFFWNLDDVEIK 180
DB 127 EEVLTAKTVVTAPEAIRGTRVQGTQKTVQLPAGTKYVAFRHFCTDFFWNLDDVEIK 186
QY 181 ANGKRADFTETPSSSTHGEPAEWTTIDADGCGWMLCLSSGQLDWLTHAGGTVNVASFS 240
DB 187 ANGKRADFTETPSSSTHGEPAEWTTIDADGCGWMLCLSSGQLDWLTHAGGTVNVASFS 246
QY 241 WNGMALNPNDYLSKDVGTATKVKYVAVNDGPPGHYAVIMISKTGNTAGDFTVVFEEPT 300
DB 247 WNGMALNPNDYLSKDVGTATKVKYVAVNDGPPGHYAVIMISKTGNTAGDFTVVFEEPT 306
QY 301 NGINKGGRFGLSTADGAKPQSVWIERTVDLPAGTKYVAFRHYNGSDLNLIYLLDDIQFT 360

```

```

Db 307 NGINKGGRFGLSTADGAKPQSVWIERTVDLPAGTKYVAFRHYNGSDLNLIYLLDDIQFT 366
QY 361 MGSPTPTDYTYTVYRDGDKIKEGLTETTFEEDGVATGNHHEYCVVEKYTAGVSPKECVNV 420
DB 367 MGSPTPTDYTYTVYRDGDKIKEGLTETTFEEDGVATGNHHEYCVVEKYTAGVSPKECVNV 426
QY 421 TVDPVQFNPVQNLGTSANGQKUTLKWADPN 450
DB 427 TVDPVQFNPVQNLGTSANGQKUTLKWADPN 456

RESULT 6
AAR96022
ID AAR96022 standard; protein; 456 AA.
XX
XX AAR96022;
XX
XX 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
XX P. gingivalis haemagglutinin hgaA Harep2 product.
DE
XX
XX Haemagglutinin; hgaA; periodontal disease; vaccine; antibody; Harep2.
XX
XX Porphyromonas gingivalis; strain 381.
XX
XX WO9617936-A2.
XX
XX 13-JUN-1996.
XX
XX 11-DEC-1995; 95WO-US016108.
XX
XX 09-DEC-1994; 94US-00353485.
XX
XX (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
XX Progulske-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
XX WPI; 1996-287181/29.
DR N-PSDB; AAT30646.
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
XX
XX Claim 4; Page 107-108; 153pp; English.
XX
XX Harep2 (AAR96022) is the product of the Harep2 repeat unit (AAT30646) of
CC the hgaA gene (AAT30654) of P. gingivalis 318. It forms part of
CC haemagglutinin hgaA (see also AAR96030). Harep2 and other hgaA repeat
CC unit products (see also AAR96021 and AAR96023-24) can be obtd. from
CC transformed host cells and used as vaccines to protect humans or animals
CC against periodontal disease. Expression in Salmonella cells allows produ.
CC of live vaccine. Harep1-4 can also be used to detect the presence of anti
CC -P. gingivalis antibodies and to raise monoclonal antibodies for
CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 456 AA;

Query Match 99.7%; Score 2436; DB 2; Length 456;
Best Local Similarity 99.8%; Pred. No. 4.2e-184;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PNPNGPTTILSFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICASSAYI 60
DB 7 PNPNGPTTILSFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICASSAYI 66
QY 61 NFEQPQNPNDYLTPELSIPNGGTLTFWVCAQDANYASHYAVYASSTGNDASNANALL 120
DB 67 NFEQPQNPNDYLTPELSIPNGGTLTFWVCAQDANYASHYAVYASSTGNDASNANALL 126
QY 121 BEVLTAKTVVTAPEAIRGTRVQGTQKTVQLPAGTKYVAFRHFCTDFFWNLDDVEIK 180

```

```

Db 127 EVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWLNLDDEIK 186
QY 181 ANGKADTFETTESSTHGEAPAEWTTIDADGGQGWLCSSQLDHLTAHGGTNVVASFS 240
Db 187 ANGKADTFETTESSTHGEAPAEWTTIDADGGQGWLCSSQLDHLTAHGGTNVVASFS 246
QY 241 WNGMALNPNDNLIISKDVTGATKVKYYAVNDGFPDGHYAVMSKTGTNAGDFTVVFEETP 300
Db 247 WNGMALNPNDNLIISKDVTGATKVKYYAVNDGFPDGHYAVMSKTGTNAGDFTVVFEETP 306
QY 301 NGINKGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
Db 307 NGINKGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
QY 361 MGSSTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGVSPKECVNV 420
Db 367 MGSSTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGVSPKECVNV 426
QY 421 TVDPVQFNPVQNLITGSAVGQKVTILKWDAPN 450
Db 427 TVDPVQFNPVQNLITGSAVGQKVTILKWDAPN 456

RESULT 7
AAW69491
ID AAW69491 standard; protein; 456 AA.
XX
AC AAW69491;
XX
DT 22-DEC-1998 (first entry)
XX
DE Haemagglutinin protein hgaA, Harep3.
XX
KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
XX
OS Porphyromonas gingivalis.
XX
PN US5824791-A.
XX
PD 20-OCT-1998.
XX
PF 11-DEC-1995; 95US-00570311.
XX
PR 08-SEP-1988; 88US-00241640.
XX
PR 25-JAN-1991; 91US-00647119.
XX
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58878.
XX
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s)).
XX
PS Claim 1; Col 133-138; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hgaA haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 456 AA;
Query Match 99.7%; Score 2436; DB 2; Length 456;
Best Local Similarity 99.8%; Pred. No. 4.2e-184;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 PNPNGTITLSESFENGIPASWKTTIDADGGNNWTTTPPGCTSPAGHNSAICASSASYI 60
Db 7 PNPNGTITLSESFENGIPASWKTTIDADGGNNWTTTPPGCTSPAGHNSAICASSASYI 66
QY 61 NPEGQNDPNVLTVELSLPNGGTLTFWCAQADANVASEHYAVYASSTGNDASNFANALL 120
Db 67 NPEGQNDPNVLTVELSLPNGGTLTFWCAQADANVASEHYAVYASSTGNDASNFANALL 126
QY 121 BEVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWLNLDDEIK 180
Db 127 BEVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWLNLDDEIK 186
QY 181 ANGKADTFETTESSTHGEAPAEWTTIDADGGQGWLCSSQLDHLTAHGGTNVVASFS 240
Db 187 ANGKADTFETTESSTHGEAPAEWTTIDADGGQGWLCSSQLDHLTAHGGTNVVASFS 246
QY 241 WNGMALNPNDNLIISKDVTGATKVKYYAVNDGFPDGHYAVMSKTGTNAGDFTVVFEETP 300
Db 247 WNGMALNPNDNLIISKDVTGATKVKYYAVNDGFPDGHYAVMSKTGTNAGDFTVVFEETP 306
QY 301 NGINKGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 360
Db 307 NGINKGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFT 366
QY 361 MGSSTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGVSPKECVNV 420
Db 367 MGSSTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGVSPKECVNV 426
QY 421 TVDPVQFNPVQNLITGSAVGQKVTILKWDAPN 450
Db 427 TVDPVQFNPVQNLITGSAVGQKVTILKWDAPN 456

RESULT 8
AAW69490
ID AAW69490 standard; protein; 456 AA.
XX
AC AAW69490;
XX
DT 22-DEC-1998 (first entry)
XX
DE Haemagglutinin protein hgaA, Harep2.
XX
KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
XX
OS Porphyromonas gingivalis.
XX
PN US5824791-A.
XX
PD 20-OCT-1998.
XX
PF 11-DEC-1995; 95US-00570311.
XX
PR 08-SEP-1988; 88US-00241640.
XX
PR 25-JAN-1991; 91US-00647119.
XX
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58877.
XX
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s)).
XX
PS Claim 1; Col 127-132; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hgaA haemagglutinin protein. The

```

CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease

XX
 SQ Sequence 456 AA;
 Query Match 99.7%; Score 2436; DB 2; Length 456;
 Best Local Similarity 99.8%; Pred. No. 4.2e-184; Mismatches 1; Indels 0; Gaps 0;
 Matches 449; Conservative 0;
 QY 1 PNPNGPTTLLSEFENGIPASWKTIDADGNGNNWTTTPPGGTSPAGHNSAICASSAYI 60
 Db PNPNGPTTLLSEFENGIPASWKTIDADGNGNNWTTTPPGGTSPAGHNSAICASSAYI 66
 QY 61 NFEQPNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 Db NFEQPNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
 QY 121 BEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNLDDVEIK 180
 Db BEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNLDDVEIK 186
 QY 181 ANGKRAADTFETTESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASFS 240
 Db ANGKRAADTFETTESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASFS 246
 QY 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDTTVVFEETP 300
 Db WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDTTVVFEETP 306
 QY 301 NGINKGGARFGLSTEADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 360
 Db NGINKGGARFGLSTEADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 366
 QY 361 MGSPTPTDYTVTVYRDGPKIKEGLTETTFBEDGVATGNHGYCVVEKVTAGVSPKCVNV 420
 Db MGSPTPTDYTVTVYRDGPKIKEGLTETTFBEDGVATGNHGYCVVEKVTAGVSPKCVNV 426
 QY 421 TVDVPQFNPQNLTGSAVQKVTWKDAFN 450
 Db TVDVPQFNPQNLTGSAVQKVTWKDAFN 456

RESULT 9
 AAR96024
 ID AAR96024 standard; protein; 439 AA.

XX
 AC AAR96024;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hagA Harep4 product.

XX
 KW Haemagglutinin; hagA; periodontal disease; vaccine; antibody; Harep4.

XX
 OS Porphyromonas gingivalis; strain 381.

XX
 PN WO9617936-A2.

XX
 PD 13-JUN-1996.

XX
 PF 11-DEC-1995; 95WO-US016108.

XX
 PR 09-DEC-1994; 94US-00353485.

XX
 XX (UYFL) UNIV FLORIDA.

XX
 PA (UABR-) UAB RES FOUND.

XX
 PI Progulskte-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX
 XX WPI; 1996-287181/29.

DR N-PSDB; AAT30648.
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 PS Claim 4; Page 114-115; 153pp; English.

XX
 CC Harep4 (AAR96024) is the product of the Harep4 repeat unit (AAT30648) of
 CC the hagA gene (AAT30654) of P. gingivalis 318. It forms part of
 CC haemagglutinin hagA (see also AAR96030). Harep4 and other hagA repeat
 CC unit products (see also AAR96021-23) can be obt'd. from transformed host
 CC cells and used as vaccines to protect humans or animals against
 CC periodontal disease. Expression in Salmonella cells allows prodn. of live
 CC vaccine. Harep4-4 can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 439 AA;

Query Match 95.0%; Score 2321; DB 2; Length 439;
 Best Local Similarity 98.4%; Pred. No. 5e-175; Mismatches 3; Indels 0; Gaps 0;
 Matches 426; Conservative 3;
 QY 1 PNPNGPTTLLSEFENGIPASWKTIDADGNGNNWTTTPPGGTSPAGHNSAICASSAYI 60
 Db PNPNGPTTLLSEFENGIPASWKTIDADGNGNNWTTTPPGGTSPAGHNSAICASSAYI 66
 QY 61 NFEQPNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 Db NFEQPNPDNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
 QY 121 BEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNLDDVEIK 180
 Db BEVLTAKTVVTAPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNLDDVEIK 186
 QY 181 ANGKRAADTFETTESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASFS 240
 Db ANGKRAADTFETTESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASFS 246

QY 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDTTVVFEETP 300
 Db WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDTTVVFEETP 306

QY 301 NGINKGGARFGLSTEADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 360
 Db NGINKGGARFGLSTEADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 366

QY 361 MGSPTPTDYTVTVYRDGPKIKEGLTETTFBEDGVATGNHGYCVVEKVTAGVSPKCVNV 420
 Db MGSPTPTDYTVTVYRDGPKIKEGLTETTFBEDGVATGNHGYCVVEKVTAGVSPKCVNV 426

QY 421 TVDVPQFNPQNLT 433

Db 427 TINPTQFNPQNLT 439

RESULT 10

AAR69492
 ID AAR69492 standard; protein; 439 AA.

XX
 AC AAR69492;

XX
 DT 22-DEC-1998 (first entry)

XX
 DE Haemagglutinin protein hagA, Harep4.

XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hagA.

XX
 OS Porphyromonas gingivalis.

XX
 PN US5824791-A.

XX
 PD 20-OCT-1998.

```
XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX
XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX WPI; 1998-582627/49.
XX DR N-PSDB; AAV58879.
XX
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly:peptide(s)).
XX
XX Claim 1; Col 139-144; 101pp; English.
XX
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX invention. This sequence represents the hgaA haemagglutinin protein. The
XX polypeptides are used to produce antibodies to organisms associated with
XX periodontal disease. The antibodies are also used in purification and
XX identification procedures. The genes and polypeptides are used as
XX vaccines against periodontal disease
XX
XX Sequence 439 AA;
XX
Query Match 95.0%; Score 2321; DB 2; Length 439;
Best Local Similarity 98.4%; Pred. No. 5e-175;
Matches 426; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 PNPNPGTTTLLSESPENGIPASWKTIADGCGNNWTTTPPGGTSFAGHNSAICASSASYI 60
Db 7 PNPNPGTTTLLSESPENGIPASWKTIADGCGNNWTTTPPGGTSFAGHNSAICASSASYI 66
QY 61 NFEQPQNDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 120
Db 67 NFEQPQNDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 126
QY 121 EEVLTAKTVTTPAIPAIRGTRVQGTWYQKTVOLPAGTKYVAFRHHGCTDFWNLDDVEIK 180
Db 127 EEVLTAKTVTTPAIPAIRGTRVQGTWYQKTVOLPAGTKYVAFRHHGCTDFWNLDDVEIK 186
QY 181 ANGKRAADFTETFEESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGHGTNVVASFS 240
Db 187 ANGKRAADFTETFEESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGHGTNVVASFS 246
QY 241 WNGMALNPNDNYLISKDVGTATKVKYKYAVNDGFFGDHYAVMI SKGTGNAGDFTVVFETP 300
Db 247 WNGMALNPNDNYLISKDVGTATKVKYKYAVNDGFFGDHYAVMI SKGTGNAGDFTVVFETP 306
QY 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFWNLDDVEIK 360
Db 307 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFWNLDDVEIK 366
QY 361 MGGSPTPTDITYTYVRDGTIKI KEGLTETTFEEDGATGNHGYCUEVKYTAGVSPKECVN 420
Db 367 MGGSPTPTDITYTYVRDGTIKI KEGLTETTFEEDGATGNHGYCUEVKYTAGVSPKECVN 426
QY 421 TVDPVQNFVQNL 433
Db 427 TINPTQNFVQNL 439
RESULT 11
AAR96033
ID AAR96033 standard; protein; 1687 AA.
XX
XX AAR96033;
XX
XX 16-OCT-2003 (revised)
```

```
DT 04-SEP-1996 (first entry)
XX
XX P. gingivalis haemagglutinin hage.
XX
XX Haemagglutinin; hage; periodontal disease; vaccine; antibody.
XX
XX Porphyromonas gingivalis; strain FDC381.
XX
XX W09617936-A2.
XX
XX 13-JUN-1996.
XX
XX 11-DEC-1995; 95WO-US016108.
XX
XX 09-DEC-1994; 94US-00353485.
XX
XX (UYFL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX
XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX WPI; 1996-287181/29.
XX DR N-PSDB; AAT30656.
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
XX vaccination against periodontal disease.
XX
XX Claim 5; Page 138-143; 153pp; English.
XX
XX P. gingivalis 381 haemagglutinin hage (AAR96033) was identified as the
XX product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
XX The haemagglutinin can be obtd. from transformed host cells and used as a
XX vaccine to protect humans or animals against periodontal disease.
XX Expression in salmonella cells allows prodn. of a live vaccine. The
XX haemagglutinin can also be used to detect the presence of anti-P.
XX gingivalis antibodies and to raise monoclonal antibodies for diagnostic
XX appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 1687 AA;
XX
Query Match 84.4%; Score 2063; DB 2; Length 1687;
Best Local Similarity 86.3%; Pred. No. 8.9e-154;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;
QY 1 PNPNPGTTTLLSESPENGIPASWKTIADGCGNNWTTTPPGGTSFAGHNSAICASSASY- 59
Db 939 PNPNPGTTTLLSESPENGIPASWKTIADGCGHGWKPGNAFG---IAGYNSGCVYSESG 995
QY 60 INFEQPQNDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 996 LGGIGVLTDPNYLITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1055
QY 120 EEVLTAKTVTTPAIPAIRGTRVQGTWYQKTVOLPAGTKYVAFRHHGCTDFWNLDDVEI 179
Db 1056 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAKTKYVAFRHHGCTDFWNLDDVEI 1113
QY 180 KANGKRAADFTETFEESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGHGTNVVASF 239
Db 1114 KANGKRAADFTETFEESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGHGTNVVASF 1173
QY 240 SWNGMALNPNDNYLISKDVGTATKVKYKYAVNDGFFGDHYAVMI SKGTGNAGDFTVVFET 299
Db 1174 SWNGMALNPNDNYLISKDVGTATKVKYKYAVNDGFFGDHYAVMI SKGTGNAGDFTVVFET 1233
QY 300 PNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFWNLDDVEI 359
Db 1234 PNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFWNLDDVEI 1293
QY 360 TMGGSPTPTDITYTYVRDGTIKI KEGLTETTFEEDGATGNHGYCUEVKYTAGVSPKECVN 419
Db 1294 TMGGSPTPTDITYTYVRDGTIKI KEGLTETTFEEDGATGNHGYCUEVKYTAGVSPKECVN 1353
QY 420 VTVDPVQNFVQNLGSAVGQKVTCLKWDAFN 450
```


Query Match 84.4%; Score 2063; DB 2; Length 1704;
 Best Local Similarity 86.3%; Pred. No. 9e-154;
 Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGPTTLLSESPENGIPASWKTIADGDNNTTTPPGGTSGFAGHNSAICASSASY- 59
 DB 956 PNPNGPTTLLSESPENGIPASWKTIADGDNNTTTPPGGTSGFAGHNSAICASSASY- 1012

QY 60 INFEGPQPNLYLTPBELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
 DB 1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1072

QY 120 LEEVLTAKTVVTAPAIKRGTRVQGTWQKTVDLPAGTKYVAFRHFQCTDFFWINLDDVEI 179
 DB 1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQCTDFFWINLDDVEI 1130

QY 180 KANGKRADFTTFSSSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASF 239
 DB 1131 KANGKRADFTTFSSSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASF 1190

QY 240 SWNGMALNPNDNYLISKDVTKYKYYAVNDGPGDHYAVMISKTGTNAGDFTVFEET 299
 DB 1191 SWNGMALNPNDNYLISKDVTKYKYYAVNDGPGDHYAVMISKTGTNAGDFTVFEET 1250

QY 300 PNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDNLNILLDDIOF 359
 DB 1251 PNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDNLNILLDDIOF 1310

QY 360 TMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKECVN 419
 DB 1311 TMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKECVN 1370

QY 420 VTVDVPQFNPVQNLTGSAVGOKVTLKWDAPN 450
 DB 1371 VTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1401

RESULT 14
 AAW34843
 ID AAW34843 standard; protein; 1704 AA.
 AC AAW34843;
 XX
 XX
 XX
 XX 03-JUN-1998 (first entry)
 DE Arg-gingipain high molecular weight prepolyprotein sequence.
 XX
 XX Arg-specific gingipain protease; gingivalis; periodontal disease;
 KW vaccine; infection.
 XX
 XX Porphyromonas gingivalis.
 OS
 FH Key Location/Qualifiers
 FT Protein 1..227
 ET /note= "precursor protein"
 XX
 XX WO9734629-A1.
 XX
 XX 25-SEP-1997.
 XX
 XX 21-MAR-1997; 97WO-US004635.
 XX
 XX 22-MAR-1996; 96US-0013945P.
 XX
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA (MORE-) MOREHOUSE SCHOOL MEDICINE.
 XX
 XX Potempa J, Travis J, Genco C;
 XX WPI; 1997-479993/44.
 DR N-PSDB; AAT93872.
 XX

PT Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
 PT useful for protecting animals and humans from gingivalis and periodontal
 PT diseases.
 XX
 XX Disclosure; Page 68-73; 95pp; English.
 ES
 XX The present sequence represents an arginine-specific protease of
 CC Porphyromonas gingivalis. The following peptides, derived from Arg- and
 CC Lys-specific high molecular weight proteases, offer protection against
 CC infection: YTYTVRDSK IKELTATTE DDGVTGNHE YCEKVTAGS VSPKVC (I);
 CC YTVPEKQNG RMIVIVAKKY (II); QLPFIFDVAC VNGDFLEFMP CFALMLRAQ (III);
 CC GFNPYQPS NUTATQOK VTLKWDASTK (IV); GMEHYCEVK YTAGVSPKVC KDVTY (V);
 CC RMFNTPEPR YTVPEKQNG (VI); TPAFEDTYK RMFNTPEPR (VII); DYTYTVRDS
 CC TKIKELTAT TEEDGVATG NMEYCVCKY TAGVSPKVC (VIII); YTYTVRDSK KIKELTATTF
 CC EEDG (IX); RDGKIKELTATTFEEDGV ATGN (X); KIKELTATTFEEDGVATGN HEY (XI)
 CC ; KWDAPNGTPN PNPEN PNPEN PNPEN PNPEN PNPEN PNPEN PNPEN PNPEN PNPEN
 CC ; (XIII). They are used in vaccines to protect animals, including humans,
 CC from gingivitis and/or periodontal diseases
 XX
 XX Sequence 1704 AA;

Query Match 84.4%; Score 2063; DB 2; Length 1704;
 Best Local Similarity 86.3%; Pred. No. 9e-154;
 Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGPTTLLSESPENGIPASWKTIADGDNNTTTPPGGTSGFAGHNSAICASSASY- 59
 DB 956 PNPNGPTTLLSESPENGIPASWKTIADGDNNTTTPPGGTSGFAGHNSAICASSASY- 1012

QY 60 INFEGPQPNLYLTPBELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
 DB 1013 LGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1072

QY 120 LEEVLTAKTVVTAPAIKRGTRVQGTWQKTVDLPAGTKYVAFRHFQCTDFFWINLDDVEI 179
 DB 1073 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQCTDFFWINLDDVEI 1130

QY 180 KANGKRADFTTFSSSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASF 239
 DB 1131 KANGKRADFTTFSSSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTNVVASF 1190

QY 240 SWNGMALNPNDNYLISKDVTKYKYYAVNDGPGDHYAVMISKTGTNAGDFTVFEET 299
 DB 1191 SWNGMALNPNDNYLISKDVTKYKYYAVNDGPGDHYAVMISKTGTNAGDFTVFEET 1250

QY 300 PNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDNLNILLDDIOF 359
 DB 1251 PNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDNLNILLDDIOF 1310

QY 360 TMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKECVN 419
 DB 1311 TMGGSPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKECVN 1370

QY 420 VTVDVPQFNPVQNLTGSAVGOKVTLKWDAPN 450
 DB 1371 VTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1401

RESULT 15
 AAY67396
 ID AAY67396 standard; protein; 1704 AA.
 XX
 XX AAY67396;
 AC
 XX
 XX 25-APR-2000 (first entry)
 DT Arg-gingipain-2 amino acid sequence.
 XX
 XX Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
 KW immunogenic component; vaccine; inflammatory response; tissue damage;
 KW periodontal disease.
 XX
 XX Porphyromonas gingivalis.
 OS

```

XX FH Key Location/Qualifiers
XX FT 229..719
XX FT /note= "Amino acids 229-719 are specifically claimed"
XX FT 720..1185
XX FT /note= "Amino acids 720-1185 are specifically claimed"
XX PN US6017532-A.
XX PD 25-JAN-2000.
XX PF 08-NOV-1994; 94US-00336308.
XX PR 10-SEP-1993; 93US-00119361.
XX PR 24-JUN-1994; 94US-00265441.
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX PI Potempa JS, Travis J;
XX XPI; 2000-136659/12.
XX DR N-PSDB; AAZ60181.
XX PT New Porphyromonas gingivalis arginine-specific protease preparation
XX PT useful for preparing vaccines against periodontal disease and for
XX PT screening for Arg-gingipain inhibitors.
XX PS Claim 1; Col 29-42; 55pp; English.
XX CC This sequence represents a Porphyromonas gingivalis arginine-specific
XX CC proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
XX CC Gingipain-2 consists of a 50kD protease component non-covalently
XX CC associated with a 44kD haemagglutinin component. The proteinase is
XX CC stimulated by glycine containing peptides and glycine analogues. It is
XX CC inhibited by cysteine protease group specific inhibitors. The protease
XX CC preparation can be used in immunogenic compositions and vaccines against
XX CC inflammatory response and tissue damage caused by P. gingivalis in
XX CC periodontal disease. It can also be used to screen for agents that
XX CC modulate Arg-gingipain proteinase activity inhibitors
XX SQ Sequence 1704 AA;

Query Match 84.4%; Score 2063; DB 3; Length 1704;
Best Local Similarity 86.3%; Pred. No. 9e-154;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGFTTTLSSFENGIPASWKTIADGDGNNWTTTPPGGCTSPAGHNSAICASSASY- 59
DB 956 PNPNGFTTTLSSFENGIPASWKTIADGDGNNWTTTPPGGCTSPAGHNSAICASSASY- 59
QY 60 INFEGQNDNYLVTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
DB 1013 LGGIGVLPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1072
QY 120 LEEVLTAKTVTAPAIRGTRVQGTWYKTVQVLPAGTKYVAPRHFGCTDFWNLDDVEI 179
DB 1073 LEEITAKG-VKSPAIRG-RIQGTWRQKTVLPAGTKYVAPRHFGCTDFWNLDDVEI 1130
QY 180 KANGRADFTETFESSHGEAPAEWTTIDADGGQGLCLSSGQDLTAHGGTNVASF 239
DB 1131 KANGRADFTETFESSHGEAPAEWTTIDADGGQGLCLSSGQDLTAHGGTNVASF 1190
QY 240 SWNGMALPDNYLISKDVTGATKYKYAVNDGFGDHYAVNMISKTGNAGDFTVVFEET 299
DB 1191 SWNGMALPDNYLISKDVTGATKYKYAVNDGFGDHYAVNMISKTGNAGDFTVVFEET 1250
QY 300 PNGINKGARFGLSTEAGAKPQSWIERTVLPAGTKYVAPRHNCSDNLVILLDDIQF 359
DB 1251 PNGINKGARFGLSTEAGAKPQSWIERTVLPAGTKYVAPRHNCSDNLVILLDDIQF 1310
QY 360 TMGGSPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVBEVKYTAGVSPKECVN 419
DB 1311 TMGGSPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVBEVKYTAGVSPKECVN 1370

```

QY 420 VTVDVPVOENPYONLTGSAVGQKVTILKWDAPN 450
 DB 1371 VTINPTQFNPKVKNLKAQPDGGDVVLKWEAPS 1401

Search completed: May 18, 2004, 11:42:40
 Job time : 47.4099 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:37:00 ; Search time 11.7493 Seconds
(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-16
Perfect score: 2443
Sequence: 1 PNPNGPTTILSEFENGIPA.....QNLGSAVGQKVLKWDAPN 450
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----------|--------------------|
| 1 | 2443 | 100.0 | 2628 | 2 T28651 | hemagglutinin A - |
| 2 | 2063 | 84.4 | 1704 | 2 A55426 | gingipain R (EC 3. |
| 3 | 2040 | 83.5 | 1526 | 2 S49763 | gingipain R (EC 3. |
| 4 | 2011.5 | 82.3 | 1732 | 2 T30836 | lysine-specific cy |
| 5 | 266 | 10.9 | 991 | 2 I40229 | arginyl endopeptid |
| 6 | 132.5 | 5.4 | 1052 | 2 A29599 | conserved hypothet |
| 7 | 132.5 | 5.4 | 1341 | 2 A83412 | hypothetical prote |
| 8 | 130.5 | 5.3 | 2468 | 2 A10789 | hypothetical prote |
| 9 | 127.5 | 5.2 | 1684 | 2 C86832 | amylase A-180 - al |
| 10 | 126 | 5.2 | 1649 | 2 F90896 | hypothetical prote |
| 11 | 125.5 | 5.1 | 5291 | 2 B75622 | hypothetical prote |
| 12 | 124.5 | 5.1 | 2817 | 2 B97033 | hypothetical prote |
| 13 | 124.5 | 5.1 | 5188 | 2 B85547 | uncharacterized pr |
| 14 | 124 | 5.1 | 1904 | 2 T13256 | probable RTX famil |
| 15 | 123.5 | 5.1 | 713 | 2 B75489 | tail-host specific |
| 16 | 122 | 5.0 | 465 | 2 A47023 | hypothetical prote |
| 17 | 121.5 | 5.0 | 455 | 2 S59077 | S-layer protein - |
| 18 | 121 | 5.0 | 1090 | 2 AH2515 | cellulose 1,4-beta |
| 19 | 121 | 5.0 | 4936 | 2 AE2254 | hypothetical prote |
| 20 | 120 | 4.9 | 908 | 2 H90375 | hypothetical prote |
| 21 | 120 | 4.9 | 1345 | 2 T30944 | hypothetical prote |
| 22 | 120 | 4.9 | 1873 | 2 T34416 | surface protein pr |
| 23 | 120 | 4.9 | 2783 | 2 AD0835 | hypothetical prote |
| 24 | 120 | 4.9 | 3624 | 2 A86685 | large repetitive p |
| 25 | 119 | 4.9 | 1441 | 2 E90696 | prophage pil prote |
| 26 | 119 | 4.9 | 1461 | 2 A85547 | hypothetical prote |
| 27 | 119 | 4.9 | 1461 | 2 D86897 | hypothetical prote |
| 28 | 118.5 | 4.9 | 926 | 2 D97316 | hypothetical prote |
| 29 | 118.5 | 4.9 | 1939 | 2 D97316 | probable S-layer p |

30 116.5 4.8 1180 2 E86719 hypothetical prote
31 116.5 4.8 1385 2 T18213 parasporal crystal
32 116 4.7 2660 2 E85822 probable invasive Z
33 116 4.7 3029 2 S76109 hypothetical prote
34 115.5 4.7 715 2 J4908 alkaline serine pr
35 115.5 4.7 1034 2 T30551 beta-galactosidase
36 115.5 4.7 1118 1 A49724 protein-tyrosine-p
37 115 4.7 821 2 AD1507 probable secreted
38 114.5 4.7 607 2 C69503 conserved hypothet
39 114.5 4.7 1248 2 C89874 autolysin [importe
40 113.5 4.6 875 2 AF0472 probable outer mem
41 113.5 4.6 938 2 AF1772 internalin-like pr
42 112.5 4.6 987 2 A64474 hypothetical prote
43 111.5 4.6 1122 2 T18346 MG1 protein precu
44 111.5 4.6 1274 2 T10729 transferrin-like p
45 111.5 4.6 2167 2 AF1489 cell wall-associat

ALIGNMENTS

RESULT 1

T28651 hemagglutinin A - Porphyromonas gingivalis
C;Species: Porphyromonas gingivalis
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Prognulske-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381 contains four]
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AAB17128.1
C;Genetics:
A;Gene: haga

Query Match 100.0%; Score 2443; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 8.7e-161;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNPNGPTTILSEFENGIPASWKTIADGNGNWNWTTTPPGGTSPAGHNSAICASSASYI 60
Db 500 PNPNGPTTILSEFENGIPASWKTIADGNGNWNWTTTPPGGTSPAGHNSAICASSASYI 559
QY 61 NFEQPQNDNLYLTPELSLPNGGTLTFWVCAQDANYASBHYAVYASSTGNDSNFANALL 120
Db 560 NFEQPQNDNLYLTPELSLPNGGTLTFWVCAQDANYASBHYAVYASSTGNDSNFANALL 619
QY 121 EEVLTKTVTAPETARGTRVQGTWYQKTVLPAGTKYVAFRFGCTDFWNLDDVEIK 180
Db 620 EEVLTKTVTAPETARGTRVQGTWYQKTVLPAGTKYVAFRFGCTDFWNLDDVEIK 679
QY 181 ANGKRADFTETPSSSTHSGAPAEWTTIDAGDQGQWMLCLSSGQDMLTAHGGNNVASFS 240
Db 680 ANGKRADFTETPSSSTHSGAPAEWTTIDAGDQGQWMLCLSSGQDMLTAHGGNNVASFS 739
QY 241 WNGMALNPNDYLSKDVATGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTVVFETP 300
Db 740 WNGMALNPNDYLSKDVATGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTVVFETP 799
QY 301 NGINKGARGFGLSTEADGAKPQSVMLERTVDLPAGTKYVAFRHYNCSDNLYLLDDIOFT 360
Db 800 NGINKGARGFGLSTEADGAKPQSVMLERTVDLPAGTKYVAFRHYNCSDNLYLLDDIOFT 859
QY 361 MGSPPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEKYTAGYSPKECVNV 420
Db 860 MGSPPTPTDYTYVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEKYTAGYSPKECVNV 919
QY 421 TVDPVQFNPNQNLGSAVGQKVLKWDAPN 450
|||||

A;Residues: 1-795, 'I', 1391-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A;Cross-references: EMBL:U75366; MID:g2182811; PID:g2182812; PIDN:AAB60809.1
R;Jewiss, J.P.; Macrinia, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in Por
A;Reference number: Z20844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DBD
A;Molecule type: DNA
A;Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A;Cross-references: EMBL:AF01059; NID:G2738802; PID:G2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from *Porphyromonas gingivalis*. Isolates
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCEI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------------|--------|--------------|---------|------------|---------|---------|-------|--------|------|-----|------|------|-----|-----|-----|----|----|------|---|---|---|---|---|---|---|---|---|---|---|----|---|-----|----|----|-----|---|---|---|---|---|---|---|---|---|---|----|---|---|----|----|------|---|------|------|---|------|
| Query Match | 82.3%; | Score | 2011.5; | DB | 2; | Length | 1732; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Best Local Similarity | 84.7%; | Pred. No. | 4e-131; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Matches | 383; | Conservative | 19; | Mismatches | 41; | Indels | 9; | Gaps | 6; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| QY | 1 | PNNP | GTTL | SESF | ENGIPAS | WKTII | DADG | NNWTTT | PPPG | TSF | AHNS | SAIC | AS | SY- | 59 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| DB | 977 | PNNP | G- | TTL | SESF | ENGIPAS | WKTII | DADG | GHG | WKG | NAPG | --- | IAG | NSG | CVY | SE | FG | 1032 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| QY | 60 | IN | FE | GP | ND | PN | YL | VT | PE | LS | LP | NG | GT | TF | W | CA | D | AN | Y | A | S | H | Y | A | S | T | G | N | D | AS | N | F | AN | AL | 119 | | | | | | | | | | | | | | | | | | | | | |
| DB | 1033 | LG | G | I | G | V | L | T | P | D | N | Y | L | I | T | P | A | L | D | L | P | N | G | S | K | L | T | F | W | Y | C | A | D | AN | Y | A | S | H | Y | A | S | T | G | N | D | AS | N | F | AN | AL | 1092 | | | | | |
| QY | 120 | L | E | E | V | L | T | A | K | T | V | T | A | P | E | A | I | R | G | T | V | O | G | T | V | O | K | T | V | O | L | P | A | G | T | K | V | A | F | R | H | F | G | C | T | F | F | M | L | N | D | D | E | V | I | 79 |
| DB | 1093 | L | E | E | T | I | A | K | - | V | R | S | P | K | A | I | R | - | R | I | O | G | T | R | Q | K | T | V | D | L | P | A | G | T | K | V | A | F | R | H | F | O | S | T | M | F | I | D | L | D | E | V | I | 1150 | | |
| QY | 180 | K | A | N | G | K | R | A | D | T | E | T | P | E | S | T | H | G | E | A | P | A | E | W | T | I | D | A | D | G | O | G | W | L | C | L | S | S | Q | L | D | L | T | A | H | G | T | N | V | A | S | F | 239 | | | |
| DB | 1151 | K | A | N | G | K | R | A | D | T | E | T | P | E | S | T | H | G | E | A | P | A | E | W | T | I | D | A | D | G | O | G | W | L | C | L | S | S | Q | L | D | L | T | A | H | G | T | N | V | A | S | F | 1210 | | | |
| QY | 240 | S | N | G | M | A | L | N | P | D | N | Y | L | I | S | K | D | V | T | G | A | T | K | V | K | Y | Y | A | N | D | G | F | P | G | D | H | V | A | M | I | S | K | T | G | N | A | G | T | V | V | F | E | ST | 299 | | |
| DB | 1211 | S | N | G | M | A | L | N | P | D | N | Y | L | I | S | K | D | V | T | G | A | T | K | V | K | Y | Y | A | N | D | G | F | P | G | D | H | V | A | M | I | S | K | T | G | N | A | G | T | V | V | F | E | ST | 1270 | | |
| QY | 300 | P | N | G | I | N | G | G | A | R | F | G | L | S | T | B | A | C | A | K | P | O | S | W | I | E | R | T | V | D | L | P | A | G | T | K | V | A | F | R | H | N | C | S | D | L | N | I | L | L | D | I | I | Q | F | 359 |
| DB | 1271 | P | N | G | I | N | G | G | A | R | F | G | L | S | T | B | A | C | A | K | P | O | S | W | I | E | R | T | V | D | L | P | A | G | T | K | V | A | F | R | H | N | C | S | D | L | N | I | L | L | D | I | I | Q | F | 1330 |
| QY | 360 | T | M | G | S | T | P | T | D | Y | T | V | V | R | D | G | T | K | I | K | E | G | L | T | E | T | T | F | F | E | D | G | V | A | T | G | N | H | E | C | V | E | K | V | T | A | G | V | S | P | K | E | C | V | N | 419 |
| DB | 1331 | T | M | G | S | T | P | T | D | Y | T | V | V | R | D | G | T | K | I | K | E | G | L | T | E | T | T | F | F | E | D | G | V | A | T | G | N | H | E | C | V | E | K | V | T | A | G | V | S | P | K | E | C | V | N | 1390 |
| QY | 420 | V | T | V | D | P | V | O | F | N | P | V | Q | N | L | T | G | S | A | V | - | G | O | K | V | T | L | K | W | A | P | 449 | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C:Species: Porphyromonas gingivalis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40229
R:Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A:Title: Structural characterization of arginylginsain, a novel arginine-specific cysteine
A:Reference number: I40229; MUID:95168894; PMID:7864651
A:Accession: I40229

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match 10.9%; Score 266; DB 2; Length 991;
Best Local Similarity 25.9%; Pred.No. 1.7e-10;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;

QY 24 TIDADGNNWTTTPPGGTSPAGHNSAICASSASVINFEQPONPDNYLVTPELSLPNGG 83
||| : : : : :
Db 595 SVNVSVDYNGAIATISANGOMF---GSVVVENTATINLTG-----LTNES 637
||| : : : : :
QY 84 TLTFPWCADQANYASEHYAYASTG--NDASFANALLEEVLTAKTVVTAPAIRGTRV 141
||| : : : : :
Db 638 TLLTLTV----VGYNKETVIKTINGENPNYPQVSN-----LTATT-----QGKV 679
||| : : : : :
QY 142 QGTWYQKTVQLPAGTKYVAFRHGCTDFWINLDV-EIKANGKRADFETFFSSHTGEA 200
||| : : : : :
Db 680 TLKWDAPSRTKNATTN-TARSVDGIRELVSSDAPELLRSQAELVLEAHDVMDGS- 737
||| : : : : :
QY 201 PAEWTTTIDADGGQGMCLCSSGOLDMLTAGHGNTNVASFVN-----GMAINPDNYLSK 255
||| : : : : :
Db 738 -GVQILLDADHDQGVIPSDTHLWPNCVSPANLFAPFYTPPENADPSCSTNMIM-- 794
||| : : : : :
QY 256 DVTGAHKVKYYAVNDGFPGBDHVAMISKTGNTWAGDFTVVEETPNGKKGARFGLSTE 315
||| : : : : :
Db 795 DGTSASNI-----PAGTY-----DFAI----- 811
||| : : : : :
QY 316 ADGAKPQS---VMI-----BRTVDLPAGTKYVAFRHYNCSDLNYILLDDIOFTMGSSP 365
||| : : : : :
Db 812 ----AAPQAKIWIAGCGPTKEDDYFEAGKY----HFLMKXGSGDGTELTISGGG- 863
||| : : : : :
QY 366 TPFDITYTVVRDGTKIKEGLTTFBEDGVATGNHEVCVEVKYTAGVSKPCVNVTVDPV 425
||| : : : : :
Db 864 --SDYTYTVVRDGTKIKEGLTETYRDAGMSAQSHCVSVKVAAGVSPKVCVDYIPDV 921
||| : : : : :
QY 426 QFNPVQN-LTGSNAVQKVTIL 444
||| : : : : :
Db 922 ADVTAORPYTLTVVGKTIIV 941
||| : : : : :

RESULT 6
AF2959 conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2959
A;Authors: R.W.; Stetubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,
I; Kaip, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:gl7741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)

C:Genetics:
 A;Gene: Atcu3276
 A;Map position: linear chromosome
 5.4%; Score 132.5; DB 2; Length 1052;
 Query Match
 Best Local Similarity 20.7%; Pred. No.0.32;
 Matches 122; Conservative 67; Mismatches 160; Indels 239; Gaps 33;
 QY 7 TTTTLESEFENGIPASWKITDADGGNNWTT-----TPPPGGTGFAGHNSAICASSAS 58
 Db 52 TWTSGEATIGRHSAGEATV--TGDGSKVTTGLVGGTSTDPGG--LAGNT----- 99

| | | | |
|----|-----|---|-----|
| Qy | 161 | FRHFGCTDFWINL-----DDVEIKANG---KRADFTTFESSTGEAPAEWTTIDADG | 212 |
| Dd | 476 | SR-----SD--LNVLGYNGSMTVEAGGAVKSRRDGYVATYGGST-----SAVTITGD | 520 |
| Qy | 213 | GQGWLCLSSGGOLDMLTAHGGT-NUVASFSWNGMALNPONYLI-----SKDVTCGA-TK | 262 |
| Dd | 521 | GSSWAMTGTFFVGJ--ASGATGNVTVS---NGGARATGVTLGDLAGASGWTITIGAGSK | 575 |
| Qy | 263 | VKYYY-----AVNDGFPG-----DHY-----AVMISKTGT----- | 287 |
| Dd | 576 | VTAIVDNGTVNCGSVVDYGFQSGSLSVVNGGSLDAYNLVGVNALGSSGAVLVSGVGS | 635 |
| Qy | 288 | -----NAGD-----FTVVFBETPNINKGARFGLSTBADGAKPOSV | 324 |
| Dd | 636 | VDGLMVVGNAGNSVEITGGASLAAPTILIAATEAGSTGVLSIGAGSGQTARSAGA---- | 690 |
| Qy | 325 | WIE--RTVDLPAGTKYVAFRHYNCSDNLNVILLDDIQ-----FTMGSPPTDYTY | 372 |
| Dd | 691 | -VEARAFAGAGNSGI VFNH--SETGTYTSADISGAGRVAEAGVTTLSGNNS----- | 740 |
| Qy | 373 | TVYRDGTKIKEGLTETT-----FEEDG-----VA | 396 |
| Dd | 741 | --YSGGTISAGMLKGFAKSPGSGI VNNAELVVDGGTLSNAISGTSFKTGDNLLI | 798 |
| Qy | 397 | TGNEYCEVKYTAGVSPKECNVTVPDPFNPQNLTGSAVGOKVTL | 444 |
| Dd | 799 | TGNTSTSGATAVSAG-----KLVSNGSLASAVSGSVGATVGGTGTI | 839 |

RESULTS 8
A83412

C;Species: *Pseudomonas aeruginosa*

C;Accession: A83412
R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizouchi, S.D.; Warren, P.; Hickey

.; Lory, S.; Olson, M.V.

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.
A;Reference number: A82950. MUID-20437337. PMID:10984042

A;Status: preliminary

A;Residues: I-2468 <STO>
A:CROSS-references: GR:AE004613. GR:AE004091. NID:G9947956. PIDN:2AC05763 1. C9

C;Genetics:

Query Match

| | | | | | | | | | |
|---------|------|--------------|-----|------------|------|--------|------|------|-----|
| Matches | 119; | Conservative | 47; | Mismatches | 182; | Indels | 189; | Gaps | 30; |
|---------|------|--------------|-----|------------|------|--------|------|------|-----|

QY 2 NPNPGTT-----TLSEFENGIPASWKTIDADGDN-----NWT 35

00 757 N160911TUDG1ATF?GSSV I LIDG--NONE IQ--QVIADGSGNWSFIFSI FLDAGI VVNAT 81Z

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040

QY 69 --DNYLVTPELSLPNGGTLTFWVCAODANYASEHYAVYASSTGNDPASNFANALLEEVLTA 126:

Db 873 GSGNWTYTPSTPIANGTVN--VVAQDA-----AGN-SSPGASVTVDSPA 916

QY 127 KIVVTAPEAIRGTRVQGIWYQKIVQLPAGTKYVAFRHFGCTDFFWINLDDVEIKANGKRA 186

[illegible][illegible]

Db 949 G-----QVTADSG-NWSFTETPLANGTVVNNATASDPTGNTSAPAS 989
Qy 239 PSWNGMA-----LNPDNYSKDVTKVKYVAVNDGFPDGHVAVMISKTG----- 286
Db 990 TVDSVAPAAVVPNSN---GAEISGTAEPGATVTLTDSGNPIGQVTDGSGNWSFTPS 1046
Qy 287 TNAGDFTVV---FEETPNKGGARFGLSTRADGAKPOSVMWERTVDLPAGTKYVAFRHY 344
Db 1047 TPLADGTVVNATADPAG-NTGGQG---STTVDAIAPAT---PTVNLNGSSLSG--- 1094
Qy 345 NCSDLNXYLLDD-----IQTMGSGPTPTDYTVTVRDGFKIKEGLTETTFEEDGVATG 398
Db 1095 TAEFGSTVLLDGNPNPIAEVADGS---GNWYTV---PSTPIANGTVVNVVAQD---ASG 1146
Qy 399 NHEYCEVYKTAGYSPKECVN-----VTVDVQVFNPNVONLTGSAGV 439
Db 1147 NSSPPATVTVDSAPPAPVNPNSGVISGTAEGATVTLTDAGNPIGQVTDGSG 1203
RESULT 9
S10789
Query Match 5.2%; Score 127.5; DB 2; Length 1684;
Best Local Similarity 20.9%; Pred. No. 1.3;
Matches 102; Conservative 53; Mismatches 184; Indels 149; Gaps 27;
C:Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C:Accession: S10789
R: Candussio, A.; Schmid, G.; Boeck, A.
A: Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an
A: Reference number: S10789; MUID: 90336627; PMID: 1696201
A: Accession: S10789
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1684 <CAN>
A: Cross-references: EMBL:X53373; NID:948305; PIDN:CAA37453.1; PID:948306
Qy 17 GIPASWKTIDADGDNWNTTPPPGGTSFAGNSAICA-SSASVINFEQPQPNLYLTP 75
Db 204 GLPRDWTNQA--QGQWHT-----HNDIMKDNKAEANWANGSD----- 241
Qy 76 ELSLPGGTLTFWVCAOD-ANYASEHYAVYASSTGNDASNFANALLBEVLTKTV-VTAP 133
Db 242 -----WIRDETAGY-----DNCGGSEQTMCIQFLPDIKTEVITGVDL 280
Qy 134 EAIRGT-RVQGTWYQKTVQVLPAGTKYVAFRFGCTDFE--WNLDDVEIKANGKRAFTE 190
Db 281 PIERNKNDQASGYEDWF-VFAEPYRQDLNIAPKOYLKITSWBEFGIDGFRVDTAK 339
Qy 191 TFSSTHGEAPAE-----WTTIDADGGQGMCLSSGGDLMLTA-----HG----- 231
Db 340 HVEIERWAEKNEAEVALQVWRENNPKPGANW-----DNFWMTAEVFGHGLKSEYFD 394
Qy 232 -GTVNVASFV-----NGMALPNVNLISKDVTKYKYYAVNDGFP 274
Db 395 FGDSVINFEFQANFNLEGLFSRYANSINTDPDNMLSYVSSHDTKL-----YSRDD--- 448
Qy 275 GDHYAVMISKTGN---AGDFTVVR-EETPNKGGARFGLSTEADGAKPOSVW---IE 327
Db 449 -----LQAGTALLLPQGVQVYGFETARPLGDDG-----SDPEQGTSSNNWANNIN 496
Qy 328 RTVDLPAGTKYVAFR-----HYNCSDLNLY-----LLDDIQFTMGSGPTPT 368
Db 497 QNV-LSHWQKLGQRNNHIAIGAGHQLSDSPYTFARTYESDDIVDEVVATGAQGTTA 555
Qy 369 DYTIVYRDGTKIKEGLT---ETTFEEDGVATGNHNEYCEVYKTAGYSPKECVNVTVDPVQ 426
Db 556 VTVEGVFEDGTVVRDAYTGDDETTVK-GTAT-----FTAGTGGIILLIENTAEPVT 604
Qy 427 FNPVQNL 434
| : : |

Db 605 NLPVTSAT 612

RESULT 10

C86822
hypothetical protein yqbK [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86822
R: Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A: Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A: Reference number: A86625; MUID: 21235186; PMID: 11337471
A: Accession: C86822
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1649 <STO>
A: Cross-references: GB:AE005176; PID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
A: Experimental source: strain IL1403
C: Genetics:
A: Gene: yqbK

Query Match 5.2%; Score 126; DB 2; Length 1649;
Best Local Similarity 20.4%; Pred. No. 1.6;
Matches 110; Conservative 56; Mismatches 189; Indels 184; Gaps 28;

Qy 40 PGGTSFAGNSAICASSASVINFEQ-PQN-----PDNY-----LVTPELSLPN 81
Db 65 PSGTTFNG-----NSVQPNRTPLPQNSVEDAIMPNSFGWQQTIANLTLPTNQVTA 117
Qy 82 GGTITFWVCAQDANYASEHYAVYASSTGNDASNF-----ANALLEVLTAKTVVTAPEA- 135
Db 118 NGLIISGYPTDAGVQQAALYAMY-SATNASTSDFGLYGTANTLSNL-LKAVVNSPAT 175
Qy 136 -----IRGTRVQGTWYQKTVQL-----PAGTKYVAFR-----HFGCTDFFWNLDDV 177
Db 176 NMTFSTLKGHAKSLTWISNPADLLTSSNSQPSGTNYTGLPSNVYGVPTIF-----RVN 231
Qy 178 EIKANGKRAFTEFFSSHTGEAPAEWT-----SWITGAPNIYGGTDSNLTNLTIGATSIAG 216
Db 232 TVASGDNVIAQGNVAFATNG-----VASFWSNGMALNPDNLYLSKDVTKG--ATKV 263
Qy 217 LCLSSGOLDWLTAGHGTNV-----VASFWSNGMALNPDNLYLSKDVTKG--ATKV 263
Db 288 -NIYGNASATISGNTHVTHIAOSSSTINSVTGGSAGTTISGNTNL---DISGATASQI 343
Qy 264 KYIYAVNDGPPGHHYAVMISKTGTNAGDFTVFEETPNG-----INKGARFG-----LS 313
Db 344 TNYGAGIG-----TSNSPVNVNGNVTYVNSTNGGARYQLYQGGTVYGNISGIY 394
Qy 314 TEADGAKPOSVMWERTVDL-----PA-----GTKYVAFRHYNCSDLNLYL 353
Db 395 NTLSGA---GGWTGATSNINGAGGPASTNGSGFQNGITSGAGNVISNSYTSSTF--- 448
Qy 354 LDDIQFTMGSGPTPTDYTVTVRDGTKIKEGLTETTFEEDGVATGNHNEYCEVYKTAG-- 411
Db 449 TQQAFTGNGAGTASQAQA-----TNSITTAAGIILYANITVYKSAFTGTA 496
Qy 412 -----VSPKE-----CVNVTVDVQVFNPNVONLTGSAGVQKVT 443
Db 497 GAVYGVGGNHDSLKISPSOMWGLSGQTLGDSAVGVT-DAKAYGQIPSTTVVNSAQKIT 554

RESULT 11

F90696
hypothetical protein EC0542 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R: Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A: Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

QY 48 ----HNSAICASSASYINFEQPQNPNLYLTPELSLPNGGTLTFWVCAQDANYASEHYA 102
::| | : | | | | | | | | | | | | ::|

Db 981 DKYFNSGIMOTGLVYV-----NPDYGF-----DNHGHLT-----GMHSINGYI 1022
QY 103 VYASSTGNDASNFANALLEVLAKTAVTVAPEALRGTRVGTWYQKTVQ-----LPAGTKY 158
Db 1023 YFEDSTGKAQGFV-----YLGKTYFNTNMTGTFVNANNNL 1060
QY 159 VAFRHGCTDFFWINLDDVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGGQWLC 218
Db 1061 YFEDNEGVMGTWNY-----NSNRYFSA-----GASVTGFTI-----DGKCYCF 1103
QY 219 LSSGQL--DMLTAHGTNNVASFWSNGMALNP-----DNYLISKDVTGAT-- 261
Db 1104 DSGALYTDVVTINGST---YGFNTDGLMGTQITRYNNGYSSYFNTYFNSDGTAKTG 1160
QY 262 -----KVKKYYAVNDG-----FPGDHVAVMISKGTGTAGDFTVVFBETPENGINK-- 305
Db 1161 FFYILNKTYFNSDGRMLQCYQYINGHYF-----APDGTQGTWITNGSSKYYL 1212
QY 306 ---GGARFGLST-----EADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLVILL 354
Db 1213 DPGSAAVTGLQTINGNKYCFDSNGILQHN-----GIFYIGNTYYG-SDNNGIML 1260
QY 355 DDIOFTMGSGPTTDITYTYVRDGTIKKEGL 385
Db 1261 TGLQLING-----YLYCFNSDGS-VKTGL 1283

RESULT 14
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Pocamoussis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-518 <STO>
A;Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
A;Genetics:
C;Gene: Z0615

Query Match 5.1%; Score 124; DB 2; Length 5188;
Best Local Similarity 20.7%; Pred. No. 10;
Matches 123; Conservative 67; Mismatches 181; Indels 222; Gaps 34;
QY 6 GTT-----TLSSFFENGIPASMKTDADGDNNTTTPPPGGTSFAGHNSAICASS--- 56
Db 965 GTTAAAGQTLTVLNNNTYQT--TVLADG---TWSNVNP-----AADLUGTASSYTV 1013
QY 57 -ASYINFEQ-PQNPDNVYV-----TPELSLPNGGTLTFWVCAQDANYASEH-YAVYASSTG 109
Db 1014 TATVSDKAGNPASADHALVVDITAPDLTINT-----VAGDDIINAIEHGQALVVSPTS 1066
QY 110 NDASNANALLEVLAKTAVTVAPEALRGTRVGTWYQKTVQIPLAGTKYVAFRHFCTDF 169
Db 1067 TGAA--AGDVTVTGLNGKNTTTLDA-----SGNW---SVGIPAA----- 1101
QY 170 FWNLDDVEIKANGK-----ADFTTFESSHGEAPAEWTTIDADG-----DGQG 215
Db 1102 -----DVITALGTSQITFASLSDRAGNSDTHD-----VTVDSGTLTINTVSGDD 1149
QY 216 WLCLS-----SGQLDMLTAHGGTNV-----ASFWSNG-----MALNP 248
Db 1150 IINAAEIVVAQTISGGVGTGAVAGNTVIVTIGGNOYNATVQSDLSNSVSVFANVLQALGN 1209
QY 249 DNYLISKDVTGATK-----VKYYAVNDGFGPDHYAVMISKGTGNAGDFTV-----V 295

Db 1210 GELTISASLTSNANTGTATHDIVIDANLPG-----LRVDTVAGDDVINSIEHTQALV 1262
QY 296 FEETPNKNGGA-----REFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNC 347
Db 1263 ITGSSSGLAAGAAITVIVINSVITYGATVLADGSGSVGPVADVATNWPAGTVNIAVSGTNTA 1322
QY 348 -----DLNVILLDDI-----QFTMGGSPTPTD----- 369
Db 1323 GTTTSISHPTVDLAAVAITNTLSTDDVINAAEKSDIQLSGTTSGEVAGQITVIFPG 1382
QY 370 --YTYTYRDRGTIKKEGLT-----ETTFEEDGVA-----TGN-----HEYCEVEKYT 409
Db 1383 KSYTTTVAADNT---WGLTIPAVDVATLPDGAANVQASVSNVAGNSTQATHASVD--- 1435
QY 410 AGVSPKECVNVTVDPOFNPQNL--TGSAY-----GQKVTLKWDAPN 450
Db 1436 -ATAP-----SVTINTIATDILNAAEAGSALTISGTSTAEGQTVTVTLNGVN 1483

RESULT 15
T13256
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T
C;Species: Lactococcus lactis phage BK5-T
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: T13256
R;Boyce, J.D.; Davidson, B.E.; Hillier, A.J.
Appl. Environ. Microbiol. 61, 4089-4098, 1995
A;Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T
A;Reference number: Z17646; MUID:96064422; PMID:8526523
A;Accession: T13256
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1904 <BOY>
A;Cross-references: EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAA98579.1

Query Match 5.1%; Score 123.5; DB 2; Length 1904;
Best Local Similarity 22.4%; Pred. No. 3;
Matches 105; Conservative 50; Mismatches 144; Indels 169; Gaps 30;
QY 2 NPNPGTTLSESFENGIPASWKTDADGDNNTTTPP-----PGTSPAG-----H 48
Db 493 NVNDGADGKTPYFH-----TAW-SYSDGT--DRFTTYPNLNLLGSKDFSGGWLVERSE 546
QY 49 NSAICASSASYINFEGPQPNVLTPELSPNGGTLTFWVCAQDANYASEHVAVYASST 108
Db 547 NDCI-YKGLTVKSEGP-----ITKQFIAPKEGIYTF-----SAIVKSS 585
QY 109 GNDASNANALLEVLAKTAVTVAPEALRGTRVGTWYQKTVQ--LPAGTK-YVAFRHF 165
Db 586 GNA-NIIRAVTLNNVTGKIV---PDKSMG--INFQWLDRDSFQVTLKAGDKIYAQYNVAG 639
QY 166 -----CTDFF-----WINLDDVE 178
Db 640 SGVLMNAGHKWEGSAAATDGLTSNYPNLNLLGSKKYTKDNPRIQSSANDGTTTVDVDF 699
QY 179 IKANGKRAADFTTFESSHGEAPAEWTTIDADGGQWLCSSGQLDMLTAHGGTNVVAS 238
Db 700 VK-NLKAGTYTWS-----GKADAPWTHDTSANGKVKGL-----WLVS----- 737
QY 239 FSNMGALNPDNVYLSKQVGTGATKVKYYAVNDGFPDHYAVMISKGTGNAGDFTV----- 294
Db 738 -TTPGLGVN-----IS---LGETVPKTVIEVPKDG---DYCVRVN-TYSGKGDIEAHKEW 783
QY 295 VFEETPNKNGKAGFGLSTEADGAKPQSVWIE-----RTVDLPAGT-KYVAFRHYNCSD 348
Db 784 DFKLEPGSI-----ATPWPFSKNEQLPFWPFSLEVTADYPSYIGQYTFIQYD--- 836
QY 349 LNVILLDDIQTMGSGPTTDITYTYVRDGTIKKEGLTETTFEEDGVA 396
Db 837 -----SAKPSDYTWLSLR-GNDGKGAT-----GKQGVA 864

Search completed: May 18, 2004, 11:47:47

Job time : 13.7493 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 8.22454 Seconds
(without alignments)
2848.981 Million cell updates/sec

Title: US-08-570-311-16

Perfect score: 2443

Sequence: 1 PNPNGTTLSEFENGIPFA.....QNLGSAVGQKVLKWDAPN 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 2443 | 100.0 | 2628 | 1 HGA2_PORGI | Q51845 porphyron |
| 2 | 2434 | 99.6 | 2164 | 1 HGA1_PORGI | P59915 porphyron |
| 3 | 1460 | 59.8 | 989 | 1 PRTH_PORGI | P46071 porphyron |
| 4 | 266 | 10.9 | 991 | 1 CPGL_PORGI | P28784 porphyron |
| 5 | 126 | 5.2 | 721 | 1 OGP_MOUSE | Q62010 mus musculus |
| 6 | 121.5 | 5.0 | 465 | 1 SLAP_LACER | Q05044 lactobacill |
| 7 | 121 | 5.0 | 1090 | 1 GUXB_CELFI | P50899 cellulomona |
| 8 | 120 | 4.9 | 2660 | 1 YEEV_ECO57 | Q8x8v7 escherichia |
| 9 | 117.5 | 4.8 | 1256 | 1 ATL_STAAR | P52081 staphylococ |
| 10 | 116.5 | 4.8 | 331 | 1 PME_ASPAC | Q12535 aspergillus |
| 11 | 116.5 | 4.8 | 1385 | 1 C5AA_BACUD | Q45760 bacillus th |
| 12 | 114.5 | 4.7 | 607 | 1 YK28_ARCFU | Q28251 archaeoglob |
| 13 | 112.5 | 4.6 | 987 | 1 YD94_MENTJ | Q58789 methanococc |
| 14 | 111.5 | 4.6 | 1122 | 1 ADPL_MYCGA | Q49379 mycoplasma |
| 15 | 110.5 | 4.5 | 1034 | 1 BGAL_BACME | Q52847 bacillus me |
| 16 | 109.5 | 4.5 | 282 | 1 PRTA_ASPNG | P24665 aspergillus |
| 17 | 109.5 | 4.5 | 1953 | 1 BIGA_SALTY | P25927 salmonella |
| 18 | 109 | 4.5 | 872 | 1 GUXA_CELFI | P50401 cellulomona |
| 19 | 108.5 | 4.4 | 1260 | 1 ALSI_CANAL | P46590 candida alb |
| 20 | 106.5 | 4.4 | 1200 | 1 HVAL_SREPU | Q76536 strongyloce |
| 21 | 106 | 4.3 | 827 | 1 XANP_XANS2 | Q60106 xanthomonas |
| 22 | 106 | 4.3 | 2812 | 1 ZAN_HUMAN | Q9y493 homo sapien |
| 23 | 105.5 | 4.3 | 524 | 1 CHID_BACCI | P27050 bacillus ci |
| 24 | 105.5 | 4.3 | 1534 | 1 YFAS_ECO57 | Q8xe35 escherichia |
| 25 | 105 | 4.3 | 1045 | 1 PRTS_SERMA | P09489 serratia ma |
| 26 | 105 | 4.3 | 1300 | 1 L20K_RICRI | P14914 rickettsia |
| 27 | 105 | 4.3 | 1654 | 1 OMPB_RICRI | Q53047 r outer mem |
| 28 | 105 | 4.3 | 2249 | 1 OMPA_RICRI | P15921 rickettsia |
| 29 | 104.5 | 4.3 | 569 | 1 YAP3_YEAST | P32329 saccharomyc |
| 30 | 104.5 | 4.3 | 1122 | 1 ADP2_MYCGA | Q9rem8 mycoplasma |
| 31 | 104.5 | 4.3 | 2124 | 1 PGCA_RAT | P07897 rattus norv |
| 32 | 104 | 4.3 | 609 | 1 HAPT_VIBCH | P24153 vibrio chol |
| 33 | 104 | 4.3 | 1045 | 1 GUNE_CELFI | P26225 cellulomona |

34 103.5 4.2 1861 1 APU_THETU P38536 t amylopull
35 103 4.2 435 1 AM3D_ORISA P27933 oryza sativ
36 103 4.2 548 1 THER_BACST P06874 bacillus st
37 103 4.2 1044 1 ITA8_CHICK P26009 gallus gall
38 103 4.2 1157 1 C8AA_BACUK Q45704 bacillus th
39 103 4.2 2725 1 FLNC_HUMAN O14315 homo sapien
40 102.5 4.2 551 1 AMVB_THETU P19584 thermoanaer
41 102.5 4.2 1220 1 CSAC_BACTU P56955 bacillus th
42 102.5 4.2 2132 1 PGCA_MOUSE Q61282 mus musculu
43 102 4.2 1289 1 CSAB_BACUD Q45753 bacillus th
44 101.5 4.2 1045 1 PRTH_SERMA P29805 serratia ma
45 101.5 4.2 1341 1 VG37_BPT2 P07067 bacterioph

ALIGNMENTS

RESULT 1
HGA2_PORGI STANDARD; PRT; 2628 AA.
AC Q51845;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin A precursor.
GN HAGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97047672; PubMed=8926061;
RA Han N., Whitlock J., Progulsk-Fox A.;
RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
RT contains four large, contiguous, direct repeats.";
RL Infect. Immun. 64:4000-4007(1996).
CC -!- FUNCTION: Agglutinates erythrocytes.
CC -!- SIMILARITY: Belongs to peptidase family C25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; U41807; AAB17128.1; -
CC PIR; T28651; T28651.
CC Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 100.0%; Score 2443; DB 1; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.1e-163;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNPNGTTLSEFENGIPASWKTTIDADGNNWTTTPPGGTSPAGHNSAICASSASYI 60
|||
Db 500 PNPNGTTLSEFENGIPASWKTTIDADGNNWTTTPPGGTSPAGHNSAICASSASYI 559
|||
QY 61 NFEQPONPNYLVTPELSLPNGGTLTFWVCAQDANYASHEHYAVYASSTGNDASNFANALL 120
|||
Db 560 NFEQPONPNYLVTPELSLPNGGTLTFWVCAQDANYASHEHYAVYASSTGNDASNFANALL 619
|||

QY 121 EVLTAKTAVTAPAIRTRVQGTWQKTVQVLPAGTKVYAFRHFCTDFWNLDDVEIK 180
 DB 620 EVLTAKTAVTAPAIRTRVQGTWQKTVQVLPAGTKVYAFRHFCTDFWNLDDVEIK 679
 QY 181 ANGKADTFETPESSTHGEAPAEWTTIDADGGQWMLCLSSQDLWLTAGGTNNVVSFS 240
 DB 680 ANGKADTFETPESSTHGEAPAEWTTIDADGGQWMLCLSSQDLWLTAGGTNNVVSFS 739
 QY 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 300
 DB 740 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 799
 QY 301 NGINKGARFGLSTEADGAKPOSVMWERTVLDLPAGTKVYAFRHFCTDFWNLDDVEIK 360
 DB 800 NGINKGARFGLSTEADGAKPOSVMWERTVLDLPAGTKVYAFRHFCTDFWNLDDVEIK 859
 QY 361 MGSSTPTDITYTVYVRDGTKEGLTETTFEEDGVATGNHGYCVVEKYTAGVSPKCVNV 420
 DB 860 MGSSTPTDITYTVYVRDGTKEGLTETTFEEDGVATGNHGYCVVEKYTAGVSPKCVNV 919
 QY 421 TVDPVQFNPVQNLGTSAGVQKVTLLKWDAPN 450
 DB 920 TVDPVQFNPVQNLGTSAGVQKVTLLKWDAPN 949

RESULT 2

HGAL PORGI
 ID PRTH PORGI STANDARD; PRT; 2164 AA.
 AC P59915;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hemagglutinin A precursor.
 GN HAGA OR PGL837.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 CC Porphyromonadaceae; Porphyromonas.
 RX NCBI_TaxID=837;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=22829867; PubMed=12949112;
 RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
 RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
 RA Dewhirst F.E., Fraser C.M.
 RT "Complete genome sequence of the oral pathogenic bacterium
 Porphyromonas gingivalis strain W83."
 RL J. Bacteriol. 185:5591-5601(2003).
 CC -!- FUNCTION: Agglutinates erythrocytes (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family C25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC TIGR; AE017178; AAQ66831.1; ALT_INIT.
 DR TIGR; PGL837; -
 KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;

Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2164 HEMAGGLUTININ A.
 FT DOMAIN 26 539 PEPTIDASE C25-LIKE 1.
 FT DOMAIN 540 991 PEPTIDASE C25-LIKE 2.
 FT DOMAIN 992 1443 PEPTIDASE C25-LIKE 3.
 SQ SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;

Query Match

99.6%; Score 2434; DB 1; Length 2164;

Best Local Similarity 99.6%; Pred. No. 3.7e-163;
 Matches 448; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PNPNGTITLSESEFENGIPASWKTIADGDCGNNWTTTTPPGGTSPAGHNSAICASSASYI 60
 DB 952 PNPNGTITLSESEFENGIPASWKTIADGDCGNNWTTTTPPGGTSPAGHNSAICASSASYI 1011
 QY 61 NFGPQNDNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASFANALL 120
 DB 1012 NFGPQNDNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASFANALL 1071
 QY 121 EVLTAKTAVTAPAIRTRVQGTWQKTVQVLPAGTKVYAFRHFCTDFWNLDDVEIK 180
 DB 1072 EVLTAKTAVTAPAIRTRVQGTWQKTVQVLPAGTKVYAFRHFCTDFWNLDDVEIK 1131
 QY 181 ANGKADTFETPESSTHGEAPAEWTTIDADGGQWMLCLSSQDLWLTAGGTNNVVSFS 240
 DB 1132 ANGKADTFETPESSTHGEAPAEWTTIDADGGQWMLCLSSQDLWLTAGGTNNVVSFS 1191
 QY 241 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 300
 DB 1192 WNGMALNPDNYLISKDVTGATKVKYYAVNDGFGPDGHYAVMISKTGTNAGDFTVVFEETP 1251
 QY 301 NGINKGARFGLSTEADGAKPOSVMWERTVLDLPAGTKVYAFRHFCTDFWNLDDVEIK 360
 DB 1252 NGINKGARFGLSTEADGAKPOSVMWERTVLDLPAGTKVYAFRHFCTDFWNLDDVEIK 1311
 QY 361 MGSSTPTDITYTVYVRDGTKEGLTETTFEEDGVATGNHGYCVVEKYTAGVSPKCVNV 420
 DB 1312 MGSSTPTDITYTVYVRDGTKEGLTETTFEEDGVATGNHGYCVVEKYTAGVSPKCVNV 1371
 QY 421 TVDPVQFNPVQNLGTSAGVQKVTLLKWDAPN 450
 DB 1372 TVDPVQFNPVQNLGTSAGVQKVTLLKWDAPN 1401

RESULT 3

PRTH PORGI
 ID PRTH PORGI STANDARD; PRT; 989 AA.
 AC P46071;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Protease prth (EC 3.4.22.-).
 GN PRTH.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 CC Porphyromonadaceae; Porphyromonas.
 RX NCBI_TaxID=837;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=95012612; PubMed=7927685;
 RA Fletcher H.M., Schenkeln H.A., Macrina F.L.;
 RT "Cloning and characterization of a new protease gene (prth) from
 RT Porphyromonas gingivalis."
 RL Infect. Immun. 62:4279-4286(1994).
 RN [2]
 RP ERRATUM.
 RA Fletcher H.M., Schenkeln H.A., Macrina F.L.;
 RL Infect. Immun. 62:5707-5707(1994).
 CC -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
 CC P. GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
 CC IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
 CC AND IS A VIRULENCE FACTOR.
 CC -!- SUBCELLULAR LOCATION: In membrane vesicles.
 CC -!- SIMILARITY: Belongs to peptidase family C25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC modified and this statement is not removed. Usage by and for commercial


```

QY 426 QENPVON-LTGSVAGOKVTL 444
Db 922 ADVTAQKPYTLTVGKTIIV 941

RESULT 5
OGP_MOUSE
ID OGP_MOUSE STANDARD; PRT; 721 AA.
AC Q62010;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
DE (Oviductin) (Astragen-dependent oviduct protein).
GN OVGPI OR OGP OR CHITS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Oviduct;
RA MEDLINE=96115001; PubMed=7492680;
RA Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,
RA Araki Y.;
RT "Molecular cloning and characterization of a mouse oviduct-specific
glycoprotein.";
RL Biol. Reprod. 53:285-294 (1995).
CC -!- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
in the fertilization process and/or early embryonic development.
CC -!- SUBCELLULAR LOCATION: Secretory granules.
CC -!- TISSUE SPECIFICITY: Epithelial cells of the oviduct.
CC -!- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; D32137; BAA06863.1; -
DR MGD; MGI:106661; Ovgp1.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE_18; FALSE NEG.
KW Glycoprotein; Fertilization; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 721
FT DOMAIN 486 632
FT CARBOHYD 402 402
FT CARBOHYD 442 442
FT CARBOHYD 469 469
FT SEQUENCE 721 AA; 78807 MW; 37246C8F01665652 CRC64;

Query Match 5.2%; Score 126; DB 1; Length 721;
Best Local Similarity 18.6%; Pred No. 0.22;
Matches 103; Conservative 53; Mismatches 178; Indels 220; Gaps 22;

QY 45 FAGHNSATCA-----SSASYINF----- 62
Db 222 FTGHSPLFLSPEDSKSAYAMWKRGLTGPADKLIMGFPPTVGRNFYLLKESKNGLOQTAS 281
QY 63 EGPQPNPNVLYTPELSPNGGTLTWVCAQDNAYASEHYAVYASSTGNDASNFANALLEE 122
Db 282 MGPASPGKY-----TKQAGFLAYEYVCSFVQRAKCHWIDYQ----- 317
QY 123 VLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINLDDV----- 177

```



```

QY 177 VEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQOQWCLSSQGLDMLTAHGNTNV 236
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
677 AQTAVAGALADYARLPDDGT-----TTPDTPAP-----TVPTG-----LQAGVVTSFE 719
QY 237 ASFSNNGMALNPENYLSKDV-TGATKVKYVAVNDGFPDGHYAVNMISKTGNAQD---F 292
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
720 ATISWT--ASTDTRVTGVDVVRGATKV-----GTATTSFTDTGLTASTAYAY 766
QY 293 TVVFEETPGINKGARFGSLTEA---DGAKPOSVMIERTVDLPGATKYVAFRHYNCSDL 349
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
767 TVRAFDAAGNVSAPSAAITVTKATSDITAP-----SVPAITS-----SSSTA 810
QY 350 NYILLDDIOFTWGGSTPTDYT-----YTVRGTGTKKEGLTFTFEEDGV-ATGNHE 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
811 NSV-----TIGWSAS--TDNAGSGGLAGYVVRGATRAVQ--TTALTFTDTGLTASTAYE 861
QY 402 YCVEVKYTAG--VSPKECVNVTV-----DPQFNPQNLGSAVCO-KVTLKWA 448
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
862 YTVRADVAGNVSAPSSTAVSVTKSDTTPDTAPSVFAGLAAMTVTETSVALTWNA 917

RESULT 8
YEEJ EC057
AC Q8XV7; Q8X2B9; Q8X2C0; PRT; 2660 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein yeej.
GN Z3135 OR ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005423; AAG57041.1; --
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.

```

```

DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big 1.
DR InterPro; IPR003355; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR000601; PKD.
DR Pfam; PF02369; Big 1; 16.
DR PRINTS; PRO1369; INTIMIN.
DR SMART; SM00634; BID 1; 16.
DR SMART; SM00089; PKD; 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738..834 BIG-1 1.
FT DOMAIN 840..929 BIG-1 2.
FT DOMAIN 931..1033 BIG-1 3.
FT DOMAIN 1042..1132 BIG-1 4.
FT DOMAIN 1134..1236 BIG-1 5.
FT DOMAIN 1245..1335 BIG-1 6.
FT DOMAIN 1337..1439 BIG-1 7.
FT DOMAIN 1448..1539 BIG-1 8.
FT DOMAIN 1548..1652 BIG-1 9.
FT DOMAIN 1653..1750 BIG-1 10.
FT DOMAIN 1751..1855 BIG-1 11.
FT DOMAIN 1856..1957 BIG-1 12.
FT DOMAIN 1963..2056 BIG-1 13.
FT DOMAIN 2065..2156 BIG-1 14.
FT DOMAIN 2157..2252 BIG-1 15.
FT DOMAIN 2254..2355 BIG-1 16.
SQ SEQUENCE 2660 AA; 280062 MW; 01BE92A08F5C09D2 CRC64;

Query Match 4.9%; Score 120; DB 1; Length 2660;
Best Local Similarity 22.5%; Pred. No. 3.2;
Matches 105; Conservative 58; Mismatches 198; Indels 106; Gaps 24;

QY 7 TTTTSESPE---NGIPASWKITDADGCGNNMTTPPPGCTSFAGHNSAICASSAYINFE 63
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1470 TATVKDQFDNEVNNLPVTFSTA-----SSGLTLPGESNTNESGTAQTLGAVF----- 1519

QY 64 GPQPNPNLYVTPELSLPGG-----TLTF---WVCAQDANYASEHYAVYASTGNDASNF 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1520 GEQ-----TVTASLANNGASDNKTVHFIGDTAAAKIILTTPVPSDIAGTPQNSSGS- 1571

QY 116 ANALLEEVLTAKTVVTAPEARTGRTVQGTWQKVQVLPAGTKYVAFRFGCTDFFWIND 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1572 -----VITATVVDNNGFPVKGVTVNFTSNAATAEMTNGQAVTNEQKATVTVNTRS 1624

QY 176 DVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQO-QWCLSSQGLDMLTAHGNTN 234
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1625 SIE---SGAPD---TVEASLENGSSTLSTINNVADASTAHLTLQLQALFTVSAGDTTN 1678

QY 235 V---VAFSWNG-----MALNPENYLSKDVATKVKYVAVNDGFPDGHYAVNMISK 284
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1679 LYIEVKYNGVGPQOEVTLSVSP-----SEGVTSPNNAIY---TTNHDGNFYA---SF 1726

QY 285 TGTNAGDPTVV-----PEETENGI-NKGGARFGLSTEADGAKPQSVMIERTVDLPAG 335
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1727 TATKAGYQVATLENGDSMQQTVTVNPNVANASISLAKDPV-----IANNNDITTL 1780

QY 336 TKYVAFRHYNC---SDLNYILLDDIQ---FTWGGSTPTDTYTVYRGTGKI---KEGLTE 387
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1781 TATVADTEGNAIANSEVFTLPEDVRANFTLG-----DGGKVVDTTEGKAK 1826

QY 388 TTFEDGVAUGHYCYEVKVTAGVSPKECVNVTVDPQFNPQNL 434
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1827 VTLC--GTKAGAH--TVTASMAGGKSEQLVNVFIADTLTAQVNLNVT 1869

RESULT 9
ATL STAAU
ID ATL STAAU STANDARD; PRT; 1256 AA.
AC P52081;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bifunctional autolysin precursor [includes: N-acetylmuramoyl-L-alanine

```


DE amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
 DE acetylglucosamidase (EC 3.2.1.96)].
 GN ATL.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 ON NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
 RC STRAIN=RM450;
 RX MEDLINE=95116542; PubMed=7816834;
 RA Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
 RA Tomasz A.;
 RT "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
 RT alanine amidase domain and an endo-beta-N-acetylglucosaminidase
 RT domain: cloning, sequence analysis, and characterization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 8325-4;
 RA Foster S.J.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ENDOLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN
 CC HIGH-MANNOSE GLYCOPOLYMER AND GLYCOPOLYMER CONTAINING THE
 CC -[(MAN)5(GLCNAc)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
 CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
 CC OLIGOSACCHARIDE IS RELEASED INTACT.
 CC -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
 CC residues and L-amino acid residues in certain bacterial cell-wall
 CC glycopeptides.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
 CC unit in high-mannose glycopeptides and glycoproteins containing
 CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
 CC residue remains attached to the protein; the rest of the
 CC oligosaccharide is released intact.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
 CC EXTRACELLULAR LYTIC ENZYMES.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
 CC ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
 CC GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D17366; BAA04185.1; -
 CC EMBL; L41499; AAA99982.1; -
 CC InterPro; IPR002502; Amidase_2.
 CC Pfam; PF01510; Amidase_2; 1.
 CC DR PFam; PF01832; Amidase_4; 1.
 CC DR SMART; SM00644; Ami2; 1.
 CC DR SMART; SM00047; Lx22; 1.
 CC Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
 KW SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 1256 BIFUNCTIONAL AUTOLYSIN.
 FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
 FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
 FT REPEAT 425 589 1.
 FT REPEAT 596 758 2.
 FT REPEAT 770 932 3.
 SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;
 Query Match 4.8%; Score 117.5; DB 1; Length 1256;
 Best Local Similarity 22.1%; Pred. No. 1.8;
 Matches 103; Conservative 47; Mismatches 163; Indels 153; Gaps 26;
 QY 1 PNPENPQT---LSBSFENGIPASMKTDADGDG---NNWTTTPP-----40

598 PTPKSTPTNNKLTVSSLGV-----AQINAKNNGLFTTYVDKTKETKEVQKTFVATKE 653
 Db
 QY 41 ---GGTSF---AGNSA-----ICASSASYINFGPON-PDNYLVTPPE---LSLPNG--- 82
 Db
 654 ASLGGNKFYLVKDYNSPTLIGWVKQGDVIYNNAKSPVNMQTYTVKPGTKLYSPWGTYK 713
 QY
 83 -----GTLTFWVCAQDANYASEHYAVYASTGNDASNFAN-----ALLEBVL 125
 Db
 714 QEAGAVSGTGNQTFKATKQQQIDKS-----LYLFTVNGKSGWVSKAYLAVPAAPKAAVAQ 769
 QY
 126 AKT-----VVTAPPAIR-----GTRVQGTWYQKTVQVLPAGTKYVAFRHFSGT-- 167
 Db
 770 PKTAVKAYTVTKQTQTVSKIAQVKNNTGIRASVVEKTA-NGAKY-ADRTFYVYTK 826
 QY
 168 -----DFFWINLDVEIKANGKRADFTE--TFSESTHGEAPAEWTTI 207
 Db
 827 RAHGNETYVLLNNTSHNIPLGWENFVKDLNVQNLGKEVKTQKVTNKNNGLSMVPWGT- 885
 QY
 208 DADGDGQWCLSSGGLDMLTAHGGTNVVASFSWN-----GMLNPDNVLISKDVTGAT 261
 Db
 886 -----KNQVILGNNI-----AGTFNATKQVSGKDVLYGTINNRTGWNNAKDLTAPT 935
 QY
 262 KVK-----YYAVNDGFFGDHYAVMISKTGT-----NAGDFTVVFEETENG--- 302
 Db
 936 AVKPTTSAADKDYNTYVIKNG-NGYYVYVTPNSDTAKYSLKAFNEQPFVAVKEQVINGQTW 994
 QY
 303 ----INKGARFGLSTEADGAK-----POSVMERTVDLPAGTKY 338
 Db
 995 YGKLSNGKLAWIKST--DLAKELIKYNTGTMNLNQVAQIQAGLQY 1038
 PME_ASAPC STANDARD; PRT; 331 AA.
 ID Q12535;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pectinesterase precursor (EC 3.1.1.11) (Pectin methylsterase) (PE).
 GN PME1.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN NCBI_TaxID=5053;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KSM 510;
 RX MEDLINE=97079238; PubMed=8920970;
 RA Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
 RA Dorreich K., Dalboege H., Kauppinen S.;
 RT "Pectin methyl esterase from Aspergillus aculeatus: expression
 RT cloning in yeast and characterization of the recombinant enzyme.";
 RL Biochem. J. 319:705-712 (1996).
 CC -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
 CC -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
 CC -!- SIMILARITY: Belongs to the pectinesterase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U49378; AAB42153.1; -
 DR InterPro; IPR000070; Pectinesterase.
 DR Pfam; PF01095; Pectinesterase_1.
 DR PROSITE; PS00800; PECTINESTERASE_1; 1.
 DR PROSITE; PS00503; PECTINESTERASE_2; 1.
 KW Hydrolase; Aspartyl esterase; Cell wall; Signal.
 FT SIGNAL 1 17 BY SIMILARITY.

DR

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND MJ1394.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000963; AAB89227.1; -;
 CC DR PIR; C69503; C69503.
 CC DR TIGR; AF2028; -;
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 16 36
 FT TRANSMEM 583 603 POTENTIAL.
 SQ SEQUENCE 607 AA; 66215 MW; F46F116B2A2AFEE1 CRC64;
 CC
 CC Query Match 4.7%; Score 114.5; DB 1; Length 607;
 CC Best Local Similarity 18.4%; Pred. No. 1.1;
 CC Matches 90; Conservative 67; Mismatches 187; Indels 145; Gaps 24;
 CC
 CC 3 PNPOTTLSEFENGIPASWKTIID-ADG-----DGNWTTTPPPCGTSFAGHN 49
 CC 201 PSSVTLTWKLNLSYDINWKLTFADGSSNPKEIISGTLNNKWLITTFKWKGNYSIDE 260
 CC
 CC 50 ---SAICASSASYNIFGQPNPDNYLVTPELSLNGGTLTFWVCAQDANYASEHYAYVAS 106
 CC 261 AFVSRVLGFGAYPQF-GPINGN-----ISGTKIIDV-----PAIGNASIGVN 302
 CC 107 STGNDASN--FANALLEVLTAKTVVTAPEAIRGT-----RVQGTWYQKTVQLPAGTKYV 159
 CC 303 KSGPDNSNQWTVNTIKNTATGLTYIVKSVKWTATDRYNEINGARYENTTV----- 354
 CC 160 AFRHFGCTDFEWINLDVEIKANGKRAFDTETFSSTHGEAPAEWTIID---ADGDGQGW 216
 CC 355 -----NVQI---GRDESFTSKDLSFYQDKVPIINGVTFELVEDANYGW 395
 CC 217 LCLSSGQDLWLTAHGNTNVASFWNGKALNPDNYLISKDVTKATKYKYVAVNDGPPGD 276
 CC 396 ---GVGQ-DKITDGNNTYIIRI-----YVIGSVLVKVK-HVESAGNDIY--- 436
 CC 277 HYAVMISKTGNTAGDFTVPEETENGKKGAREGLST-EADGAKPO---SVMLERTVDL 332
 CC 437 NITLVENLGGQSPYVTVDLIPK-----NFSLNGNDWKDPQDRDGMWKKSSML 489
 CC 333 PAGTKYVAFRHYNGSLNLYLLDDIQFTMGSGPTPTDYTVV-----YRDGK 380
 CC 490 AGGPETIT-----NIQL-----SGYDTGYWWRIRPINASADGDGAYDDYTE 530
 CC 381 IKESGLIETTFEEDGATGNHCVKVTAGVSPKCVNVTVDP-----VQNPV 430
 CC 531 IENNQTVVIFYQ---IQGSEDKLLDAFIVGIDPILSMNEQTSPTKITLVSGAKATSYESV 587
 CC 431 QNLTGSAVG 439
 CC 588 MALATALVG 596
 CC
 CC RESULT 13
 CC YD94_METJA
 CC ID_YD94_METJA STANDARD; PRT; 987 AA.

Q58789;
 15-JUL-1998 (Rel. 36, Created)
 15-JUL-1998 (Rel. 36, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein MJ1394.
 GN MJ1394.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 CX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=9689087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervatage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*.";
 RL Science 273:11058-11073(1996).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U67579; AAB99404.1; -;
 CC DR PIR; A64474; A64474.
 CC DR TIGR; MJ1394; -;
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 12 32
 FT TRANSMEM 958 978 POTENTIAL.
 SQ SEQUENCE 987 AA; 112360 MW; D1E628F8F28CA86D CRC64;
 CC
 CC Query Match 4.6%; Score 112.5; DB 1; Length 987;
 CC Best Local Similarity 19.5%; Pred. No. 3;
 CC Matches 78; Conservative 69; Mismatches 143; Indels 111; Gaps 20;
 CC
 CC 7 TTLSBSFENGIPASWKTIIDADGDNWTTTPPPGG-----TSFAGHNSAICAS 55
 CC 425 STSLNNDPENGKYLKPKKEIGRDVVVISGWVYRPSNWGGPIDRIGLEDENFDGYSFVNHY 484
 CC
 CC 56 SASVINPEGQPNPDNYLVTPELSLNGGTLTFWVCAQDANYASEHYAVASGNDASNF 115
 CC 485 S-NYISLDRETNNGNPTETISPEV-----YWNPPEDWEYYPE-LKIYSNGTITESTYY 533
 CC 116 ANALLEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHYGCTDFEWINLD 175
 CC 534 QNGSL-----AAVTST-----IDNT-YTFEDRVVHGGVYVY-----VD 566
 CC 176 DVEIKANGKRAFDTETFSSTHGEAPAEWTIIDADGQGWLCLSSGQDLWLTAHGNTNV 235
 CC 567 DLEV--NSKNFDE-----YGDKNWKYLEITSANSSEGTAVLFDG--DYFKKDYNTSN 614
 CC 236 VASFSWNGKALNPDNYLISKDVTKATKY-----KYIYAVND-----GPPGDHYAVMISK 284
 CC 615 LNAINWNTITLWNSN-----DSATLVFNVLGNYSYSDRNLIAKYGFA---KILFNY 663
 CC 285 TGTNAGDFTVPEETENGKKGAREGLST-EADGAKPOQSVMIERTVDLPAGTKYVAFRH- 343
 CC 664 NGTNT-----NTSIKGVYASG--YSITDHTGTTGEINIWIEN-----VTFKND 705
 CC 344 -----YNCSDLYILLDDIQFTMGSGPTPTDYTVTVYRDGT 379

FNKSIWIDGS-----FNKSIWIDGS

| | | | | | |
|----|----|----|----------------------------|--|------------------------------------|
| 36 | QY | 36 | TTPPGGTSGFAGHNSAI | CASSASYINPEGPQNDNYLVTPELSPLNGGTLITWVCAQDAN | 95 |
| 37 | QY | 37 | TAAPAGNTKVGYPYILLSA | ISFDATR-----NGLAALAPALGQDVG | 479 |
| 38 | Db | 38 | YASBHYA----- | YASST-----GNDASNFAN-----ALLEVLTAKTV | 129 |
| 39 | QY | 39 | Y-----HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT | 536 |
| 40 | Db | 40 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 41 | QY | 41 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 42 | Db | 42 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 43 | QY | 43 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 44 | Db | 44 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 45 | QY | 45 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 46 | Db | 46 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 47 | QY | 47 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 48 | Db | 48 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 49 | QY | 49 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 50 | Db | 50 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 51 | QY | 51 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 52 | Db | 52 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 53 | QY | 53 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 54 | Db | 54 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 55 | QY | 55 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 56 | Db | 56 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 57 | QY | 57 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 58 | Db | 58 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 59 | QY | 59 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 60 | Db | 60 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 61 | QY | 61 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 62 | Db | 62 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 63 | QY | 63 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 64 | Db | 64 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 65 | QY | 65 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 66 | Db | 66 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 67 | QY | 67 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 68 | Db | 68 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 69 | QY | 69 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 70 | Db | 70 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 71 | QY | 71 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 72 | Db | 72 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 73 | QY | 73 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 74 | Db | 74 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 75 | QY | 75 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 76 | Db | 76 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 77 | QY | 77 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 78 | Db | 78 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 79 | QY | 79 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 80 | Db | 80 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 81 | QY | 81 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 82 | Db | 82 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 83 | QY | 83 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 84 | Db | 84 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 85 | QY | 85 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 86 | Db | 86 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 87 | QY | 87 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 88 | Db | 88 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 89 | QY | 89 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 90 | Db | 90 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 91 | QY | 91 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 92 | Db | 92 | Y----- | HFVPTLAVGVSVPRGANGNI | FLGSAITWGTGNGFLODKWHSPIAVIEDAPTFTT |
| 93 | QY | 93 | Y | | |

```

RESULT 15
BGAL_BACME          STANDARD;      PERT;  1034 AA.
IID BGAL_BACME      O52847;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Beta-galactosidase (EC 3.2.1.23) (Lactase).
GN BGAW.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OCC NCBI_TaxID=1404;
OCC [1]
OCC RN SEQUENCE FROM N.A.
OCC RP STRAIN=DSM 319;
OCC RA Strey J.;
OCC RRL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -! SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ita-
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ000733; CAA04267.1; -.
CC FIR; T30574; T30574.
CC HSSP; P00722; 1BLG.
CC InterPro; IPR008979; Gal bind like

```

Search completed: May 18, 2004, 11:43:32
Job time : 10.4245 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:36:20 ; Search time 32.8982 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-16

Perfect score: 2443

Sequence: 1 PNPNGTTTSESFENGIPA.....QLTGSVGGQVKTLKWDAPN 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_arChaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | ID | Description |
|------------|--------|---------|--------------|----|-------------|
| 1 | 2064 | 84.5 | 1723 | 2 | P72197 |
| 2 | 2063 | 84.4 | 1687 | 2 | Q9R9B7 |
| 3 | 2063 | 84.4 | 1704 | 2 | Q51816 |
| 4 | 2051 | 84.0 | 1706 | 2 | Q51839 |
| 5 | 2049 | 83.9 | 1723 | 2 | P72194 |
| 6 | 2041 | 83.5 | 1358 | 2 | P96967 |
| 7 | 2040 | 83.5 | 1706 | 2 | Q51838 |
| 8 | 2031.5 | 83.2 | 1223 | 2 | Q9ZNB5 |
| 9 | 2016.5 | 82.5 | 1097 | 2 | P72196 |
| 10 | 2016.5 | 82.5 | 1732 | 2 | O07442 |
| 11 | 2011.5 | 82.3 | 1732 | 2 | Q51817 |
| 12 | 1997.5 | 81.8 | 1732 | 2 | O52050 |
| 13 | 683 | 28.0 | 925 | 2 | Q9F4J0 |
| 14 | 561.5 | 23.0 | 312 | 2 | Q9KIB3 |
| 15 | 233 | 9.5 | 293 | 2 | Q9XB9 |
| 16 | 152.5 | 6.2 | 1742 | 16 | Q9P377 |

17 140 5.7 3346 16 Q7WN54
18 135 5.5 2215 16 Q7WEN0
19 135 5.5 3988 17 Q8TEZ1
20 134 5.5 1541 16 Q7UQU9
21 133 5.4 2219 16 Q88W19
22 132.5 5.4 1115 4 Q9HD43
23 132.5 5.4 1341 16 Q8UAU1
24 130.5 5.3 2468 16 Q9I2M3
25 129.5 5.3 756 9 Q858B6
26 129.5 5.3 1386 17 Q8TI72
27 129.5 5.3 2656 5 Q9GNU3
28 129 5.3 7716 16 Q7UWZ8
29 128.5 5.3 680 2 Q52644
30 128.5 5.3 2230 16 Q7U7J7
31 128.5 5.3 2768 16 Q8E9G6
32 128 5.2 1672 16 Q8Y366
33 127.5 5.2 1684 2 Q03658
34 127 5.2 1800 2 Q9L948
35 127 5.2 8682 16 Q88RG2
36 126.5 5.2 1095 2 Q85151
37 126.5 5.2 3064 16 Q82XT8
38 126 5.2 1608 17 Q8PVI0
39 126 5.2 1649 16 Q9CFA2
40 126 5.2 2522 16 Q8EKA6
41 125.5 5.1 555 16 Q826W1
42 125.5 5.1 955 17 Q8TQ91
43 125.5 5.1 5291 16 Q8X2T1
44 124.5 5.1 691 16 Q9RZS7
45 124.5 5.1 1357 17 Q8THC8

ALIGNMENTS

RESULT 1

P72197 ID P72197 PRELIMINARY; PRT; 1723 AA.
AC P72197;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lys-gingipain.
GN KGP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RA Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT "Molecular cloning and characterization of Porphyromonas gingivalis
RT Lys-gingipain.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U54691; AAA99810.1; --
DR MEROPS; C25.002; --
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (NTP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1723 AA; 186831 MW; 4508A7B50197CEBD CRC64;
Query Match 84.5%; Score 2064; DB 2; Length 1723;
Best Local Similarity 86.3%; Pred. No. 3.1e-127;

Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGPTTILSEFENGIPASWKTIIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 DB 975 PNPNGPTTILSEFENGIPASWKTIIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 QY 60 INFEGPQNPNDYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
 DB 1032 LGGIGVLTDPNLYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1091
 QY 120 LEEVLTAKTVTVAPEAIRGTRVQGTWQVKTQVLPAGTKYVAFRHFCTDFFWNLDDVEI 179
 DB 1092 LEEVITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSDMEYIDLDEVEI 1149
 QY 180 KANGKRADETFETPESSTHGEAPAEWTTIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 DB 1150 KANGKRADETFETPESSTHGEAPAEWTTIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 QY 240 SWNGMALNPNDYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
 DB 1210 SWNGMALNPNDYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1091
 QY 300 PNGINKGAREFGLSTEADGAKPQSVWIERVTDLPAQTKYVAFRHFCTDFFWNLDDVEI 179
 DB 1270 PNGINKGAREFGLSTEADGAKPQSVWIERVTDLPAQTKYVAFRHFQSDMEYIDLDEVEI 1149
 QY 360 TMGSPPTDYYTYVYRDGTKIKESGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCEVN 419
 DB 1330 TMGSPPTDYYTYVYRDGTKIKESGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCEVN 419
 QY 420 VTVDVPQNPVQNLGSAVGQKVTLKWDAPN 450
 DB 1390 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1420

RESULT 2

Q9R9B7 PRELIMINARY; PRT; 1687 AA.

AC Q9R9B7
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hemagglutinin/protease.
 GN HAGE.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RA Han N., Dong H., Progliske-Fox A.;
 RT "Cloning and characterization of hage from P. gingivalis 381.";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF026946; AAD01810.1; -;
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25.C.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_1; 1.
 DR Protease.
 KW PROTEASE.
 SQ SEQUENCE. 1687 AA; 183702 MW; D085B516A399FE70 CRC64;

Query Match. 84.4%; Score 2063; DB 2; Length 1687;
 Best local similarity 86.3%; Pred. No. 3.5e-127;
 Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGPTTILSEFENGIPASWKTIIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 DB 939 PNPNGPTTILSEFENGIPASWKTIIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 QY 60 INFEGPQNPNDYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
 DB 996 LGGIGVLTDPNLYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1055
 QY 120 LEEVLTAKTVTVAPEAIRGTRVQGTWQVKTQVLPAGTKYVAFRHFCTDFFWNLDDVEI 179
 DB 1056 LEEVITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSDMEYIDLDEVEI 1113
 QY 180 KANGKRADETFETPESSTHGEAPAEWTTIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 DB 1114 KANGKRADETFETPESSTHGEAPAEWTTIDADGQGNWTTTPPGGTSPAGHNSAICASSASY- 59
 QY 240 SWNGMALNPNDYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
 DB 1174 SWNGMALNPNDYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1055
 QY 300 PNGINKGAREFGLSTEADGAKPQSVWIERVTDLPAQTKYVAFRHFCTDFFWNLDDVEI 179
 DB 1234 PNGINKGAREFGLSTEADGAKPQSVWIERVTDLPAQTKYVAFRHFQSDMEYIDLDEVEI 1113
 QY 360 TMGSPPTDYYTYVYRDGTKIKESGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCEVN 419
 DB 1294 TMGSPPTDYYTYVYRDGTKIKESGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCEVN 419
 QY 420 VTVDVPQNPVQNLGSAVGQKVTLKWDAPN 450
 DB 1354 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1384

RESULT 3

Q51816 PRELIMINARY; PRT; 1704 AA.

AC Q51816
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arg-gingipain-1 proteinase.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95138080; PubMed=7836351;
 RA Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
 RA Travis J., Barr P.J.;
 RT "Molecular cloning and structural characterization of the Arg-
 RT gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
 RT proteinase-adhesin polypeptide."
 RL J. Biol. Chem. 270:1007-1010(1995).
 DR EMBL; U15282; AAA69539.1; -;
 DR PIR; A55426; A55426.
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.

```

DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
FT CHAIN 228 719
NATURE 50-KDA CYSTEINE PROTEINASE
GLINGIPAIN.
SQ SEQUENCE 1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;

Query Match 84.4%; Score 2063; DB 2; Length 1704;
Best Local Similarity 86.3%; Pred. No. 3.5e-127;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

Qy 1 PNPNGPTTILSESPENGIPASWKTIADGQGNWTTTPPGTSPAGHNSAICASSASY- 59
Db 956 PNPNGPTTILSESPENGIPASWKTIADGQGNWTTTPPGTSPAGHNSAICASSASY- 59
Qy 60 INFEGPQNPNLYLTPVPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 1013 LGGIGVLTDPNLYLTPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1072
Qy 120 LEEVLTAKTAVTAPETARGTRVQGTWYQKTVQLPAGTKYVAFRHRFGCTDFFWINLDDVEI 179
Db 1073 LEEITITAGK-VRSPEARIG-RIQGTWRQKTVDLPAKTKYVAFRHRFQSTDMFYIDLDEVEI 1130
Qy 180 KANGKRADEFTTFSSSTHGEAPAEWTTTIDAGDGGQGNWCLSSGQLDWLTAGHGTNNVASF 239
Db 1131 KANGKRADEFTTFSSSTHGEAPAEWTTTIDAGDGGQGNWCLSSGQLDWLTAGHGTNNVASF 1190
Qy 240 SWNGMALNPDPNLYLSKDVTKATKYYAVVNDGPGDHYAVMISKTGNTAGDFTVVFEE 299
Db 1191 SWNGMALNPDPNLYLSKDVTKATKYYAVVNDGPGDHYAVMISKTGNTAGDFTVVFEE 1250
Qy 300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAKTKYVAFRHYNCSDNLVILLDDIOF 359
Db 1251 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAKTKYVAFRHYNCSDNLVILLDDIOF 1310
Qy 360 TMGSPPTDITYTYRDRGTIKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSPKECVN 419
Db 1311 TMGSPPTDITYTYRDRGTIKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSPKECVN 1370
Qy 420 VTVDVQFNPVQNLTGSAVGQKVLKWDAPN 450
Db 1371 VTINPTQFNPVQNLKAQPDGDDVVLKWEAPS 1401

RESULT 4
Q51839 PRELIMINARY; PRT; 1706 AA.
AC Q51839; Q51840;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Arginine-specific thiol protease precursor.
GN PRTR.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=95160709; PubMed=7857299;
RA Kirszbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA Reynolds E.C.;
RT "Complete nucleotide sequence of a gene prTR of Porphyromonas
RT gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT specific thiol endopeptidase domain and a haemagglutinin domain.";
RL Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=96311339; PubMed=8713096;
RA Slakeski N., Cleal S.M., Reynolds E.C.;

```

```

RT "Characterization of a Porphyromonas gingivalis gene prTR that encodes
RT an arginine-specific thiol proteinase and multiple adhesins.";
RL Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Reynolds E.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Slakeski N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L26341; AAC18876.1; -.
DR HSSP; P95493; 1CVR.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; P:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005336; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW Protease; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 228 1706 ARGinine-SPECIFIC THIOI PROTEASE.
SQ SEQUENCE 1706 AA; 185626 MW; E8BDF07C9813B844 CRC64;

Query Match 84.0%; Score 2051; DB 2; Length 1706;
Best Local Similarity 85.8%; Pred. No. 2.2e-126;
Matches 387; Conservative 18; Mismatches 40; Indels 6; Gaps 4;

Qy 1 PNPNGPTTILSESPENGIPASWKTIADGQGNWTTTPPGTSPAGHNSAICASSASY- 59
Db 958 PNPNGPTTILSESPENGIPASWKTIADGQGNWTTTPPGTSPAGHNSAICASSASY- 59
Qy 60 INFEGPQNPNLYLTPVPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 1015 LGGIGVLTDPNLYLTPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1074
Qy 120 LEEVLTAKTAVTAPETARGTRVQGTWYQKTVQLPAGTKYVAFRHRFGCTDFFWINLDDVEI 179
Db 1075 LEEITITAGK-VRSPEARIG-RIQGTWRQKTVDLPAKTKYVAFRHRFQSTDMFYIDLDEVEI 1132
Qy 180 KANGKRADEFTTFSSSTHGEAPAEWTTTIDAGDGGQGNWCLSSGQLDWLTAGHGTNNVASF 239
Db 1133 KANGKRADEFTTFSSSTHGEAPAEWTTTIDAGDGGQGNWCLSSGQLDWLTAGHGTNNVASF 1192
Qy 240 SWNGMALNPDPNLYLSKDVTKATKYYAVVNDGPGDHYAVMISKTGNTAGDFTVVFEE 299
Db 1193 SWNGMALNPDPNLYLSKDVTKATKYYAVVNDGPGDHYAVMISKTGNTAGDFTVVFEE 1252
Qy 300 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAKTKYVAFRHYNCSDNLVILLDDIOF 359
Db 1253 PNGINKGGARFGLSTEADGAKPQSVWIERTVDLPAKTKYVAFRHYNCSDNLVILLDDIOF 1312
Qy 360 TMGSPPTDITYTYRDRGTIKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSPKECVN 419
Db 1313 TMGSPPTDITYTYRDRGTIKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSPKECVN 1372
Qy 420 VTVDVQFNPVQNLTGSAVGQKVLKWDAPN 450
Db 1373 VTINPTQFNPVQNLKAQPDGDDVVLKWEAPS 1403

RESULT 5
P72194

```


| | | | |
|--|---|--|---------------|
| Db | 1133 | KANGKRADETTETPESSTHGEATAEWTTTIDADGGQGWLCILSSQQLDWLTARGGTNNVVSF | 1192 |
| QY | 240 | SNMGALNPNNYLISKDVTGATKKVYYAVNDGFGPDGHYAVMI SKTGTAAGDFTVVPEET | 299 |
| Db | 1193 | SNMGALNPNNYLISKDVTGATKKVYYAVNDGFGPDGHYAVMI SKTGTAAGDFTVVPEET | 1252 |
| QY | 300 | PNKINKGARFGGLSTBEADGAKPOSVMIERTVDLPAGTKVAFPHYNCSDLNYILLDDIOF | 359 |
| Db | 1253 | PNKINKGARFGGLSTBEADGAKPOSVMIERTVDLPAGTKVAFPHYNCSDLNYILLDDIOF | 1312 |
| QY | 360 | TMGSGSTPTDYYTVYVRDGTKKKEGTTETTFEBDGVATGNHHEYCVKYTAGVSPKECVN | 419 |
| Db | 1313 | TMGSGSTPTDYYTVYVRDGTKKKEGTTETTFEBDGVATGNHHEYCVKYTAGVSPKECVN | 1372 |
| QY | 420 | VTVDPPVQENPVQNLTCGSAVGQKVTLKWDAPN | 450 |
| Db | 1373 | VTVNSTQFNPVKNLKAQPDGDDVVLKWEAPS | 1403 |
| RESULT 8 | | | |
| Q9ZNB5 | | | |
| ID | Q9ZNB5 | PRELIMINARY; | PRT; 1223 AA. |
| AC | Q9ZNB5; | | |
| DT | 01-MAY-1999 | (TrEMBLrel. 10, Created) | |
| DT | 01-MAY-1999 | (TrEMBLrel. 10, Last sequence update) | |
| DT | 01-OCT-2003 | (TrEMBLrel. 25, Last annotation update) | |
| DE | 130k-RMGD | (Fragment) | |
| OS | Porphyromonas gingivalis | (Bacteroides gingivalis). | |
| OC | Bacteria; Bacteroidetes; Bacteroides | (class); Bacteroidales; | |
| OC | Porphyromonadaceae; Porphyromonas. | | |
| OX | NCBI_TaxID=837; | | |
| RN | [1] | SEQUENCE FROM N.A. | |
| RP | STRAIN=381; | | |
| RC | MEDLINE=99143166; PubMed=9988746; | | |
| RA | Shibata Y., Hayakawa M., Tagiguchi H., Shiroza T., Abiko Y.; | | |
| RT | "Determination and characterization of the hemagglutinin-associated | | |
| RT | short motifs found in Porphyromonas gingivalis multiple gene | | |
| RT | products."; | | |
| RL | J. Biol. Chem. 274:5012-5020(1999). | | |
| DR | EMBL; AB019363; BAA34341.1; -. | | |
| DR | GO; GO:0005224; F:ATP binding; IEA. | | |
| DR | GO; GO:0008234; F:cysteine-type peptidase activity; IEA. | | |
| DR | GO; GO:0003910; P:DNA ligase (ATP) activity; IEA. | | |
| DR | GO; GO:0006310; P:DNA recombination; IEA. | | |
| DR | GO; GO:0008281; P:DNA repair; IEA. | | |
| DR | GO; GO:0006260; P:DNA replication; IEA. | | |
| DR | GO; GO:0006508; P:proteolysis and peptidolysis; IEA. | | |
| DR | InterPro; IPR000977; DNA ligase. | | |
| DR | InterPro; IPR001769; Peptidase C25. | | |
| DR | Pfam; PF01364; Peptidase_C25; 1. | | |
| DR | Pfam; PF03785; Peptidase_C25_C; 1. | | |
| DR | PROSITE; PS00697; DNA_LIGASE_A1; 1. | | |
| FT | NON TER | | |
| QY | 1 | | |
| Db | 1223 AA; | 131542 MW; 00225CD2BA9F91B3 | CRK64; |
| Query Match | | | |
| Best Local Similarity 83.2%; Score 2031.5; DB 2; Length 1223; | | | |
| Matches 387; Conservative 15; Mismatches 41; Indels 9; Gaps 6; | | | |
| QY | 1 | PNPNPGTTTSLSFENGIPASMKTTIDADGDNWTTTTPPGGTSFAGHNSAICASSASY- | 59 |
| Db | 468 | PNPNPG-TTSLSFENGIPASMKTTIDADGDNWTTTTPPGGTSFAGHNSAICASSASY- | 59 |
| QY | 60 | INFEGSPNDLVTPPELSLPGGTLTFWCAQADANYASEHYAVYASSTGNDASNFANAL | 119 |
| Db | 524 | LGGI GVLTPDNYLIPALDLPNGKLTFFWCAQADANYASEHYAVYASSTGNDASNFANAL | 583 |
| QY | 120 | LREVLTAKTVTVAPEAIRTRVOGTWYQKTVLPAGTKVYAFRHFGCTDFFMINLDDVEI | 179 |
| Db | 584 | LREITITAKG-VRSPPAIRG-RQGTRWQKTVLPAGTKVYAFRHFGCTDFFMINLDDVEI | 641 |

```

QY 180 KANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLDMLTAHGGTNVASF 239
DB 642 KANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLDMLTAHGGTNVASF 701
QY 240 SWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGNAGDFTVVFEE 299
DB 702 SWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGNAGDFTVVFEE 761
QY 300 PNGINKGARFGLSTEADGAKPOSVMWERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOF 359
DB 762 PNGINKGARFGLSTEADGAKPOSVMWERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOF 821
QY 360 TMGGSPTPTDYTYVYRDGTIKIKEGLTETTFEEDGVAIGNHEYCVVEKYTAGVSPKVCN 419
DB 822 TMGGSPTPTDYTYVYRDGTIKIKEGLTETTFEEDGVAIGNHEYCVVEKYTAGVSPKVCN 881
QY 420 VTVPVQFNPNVQNLGSAV--GQKVTWKDAP 449
DB 882 VTINPTQFNPNVQNLGSAV--GQKVTWKDAP 913

RESULT 9
P72196 PRELIMINARY; PRT; 1097 AA.
AC P72196;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE TonB-linked adhesin precursor.
CN TLA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
FN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97386416; PubMed=9244265;
RA Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA Curtis M.A.;
RT "The tla gene of Porphyromonas gingivalis W50: a homologue of the
RT arginine-specific protease precursor (PpPRI) which shares sequence
RT similarity to TonB-linked receptors."
RL J. Bacteriol. 179:4778-4788(1997).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Y07618; CAA68897.1; -
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR000531; TonB_box.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 53 POTENTIAL
SQ SEQUENCE 1097 AA; 118731 MW; 73BBA337B421F8B9 CRC64;

Query Match 82.5%; Score 2016.5; DB 2; Length 1097;
Best Local Similarity 85.0%; Pred. No. 2.3e-124;
Matches 384; Conservative 18; Mismatches 41; Indels 9; Gaps 6;

QY 1 PNPNGPQTLLSEFENGIPASWKTIADGGNNWTTTTPPGGTGFAGHNSAICASSASY- 59
DB 342 PNPNGPQTLLSEFENGIPASWKTIADGGNNWTTTTPPGGTGFAGHNSAICASSASY- 59
QY 60 INFNGPQPNPNYLVTPELSLNGTGLTFWVCAQDANYASEHYAVYASSTGNDASNFAL 119

```

```

DB 398 LGGIGVLTDPNLYITPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNFAL 457
QY 120 LEEVLTAKTVTATPEARTVQGTWTKTQVLTQAGTKYVAFRHFGGTDFWMLDDVEI 179
DB 458 LEEVITATAG-VRSKPAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEI 515
QY 180 KANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLDMLTAHGGTNVASF 239
DB 516 KANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLDMLTAHGGTNVASF 575
QY 240 SWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGNAGDFTVVFEE 299
DB 576 SWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGNAGDFTVVFEE 635
QY 300 PNGINKGARFGLSTEADGAKPOSVMWERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOF 359
DB 636 PNGINKGARFGLSTEADGAKPOSVMWERTVDLPAGTKYVAFRHYNCSDNLVILLDDIOF 695
QY 360 TMGGSPTPTDYTYVYRDGTIKIKEGLTETTFEEDGVAIGNHEYCVVEKYTAGVSPKVCN 419
DB 696 TMGGSPTPTDYTYVYRDGTIKIKEGLTETTFEEDGVAIGNHEYCVVEKYTAGVSPKVCN 755
QY 420 VTVPVQFNPNVQNLGSAV--GQKVTWKDAP 449
DB 756 VTINPTQFNPNVQNLGSAV--GQKVTWKDAP 787

RESULT 10
O07442 PRELIMINARY; PRT; 1732 AA.
AC O07442;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Lysine-specific cysteine proteinase.
CN PRTK.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
FN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99235907; PubMed=10219167;
RA Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RA "Characterization of a Porphyromonas gingivalis gene prtK that encodes
RA a lysine-specific cysteine proteinase and three sequence-related
RA adhesins."
RL Oral Microbiol. Immunol. 14:92-97(1999).
DR EMBL; U75366; AAB60809.1; -
DR MEROPS; C25.002; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25_1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1732 AA; 187914 MW; 45DSB91377391703 CRC64;

Query Match 82.5%; Score 2016.5; DB 2; Length 1732;
Best Local Similarity 85.0%; Pred. No. 4.2e-124;
Matches 384; Conservative 18; Mismatches 41; Indels 9; Gaps 6;

QY 1 PNPNGPQTLLSEFENGIPASWKTIADGGNNWTTTTPPGGTGFAGHNSAICASSASY- 59
DB 977 PNPNGPQTLLSEFENGIPASWKTIADGGNNWTTTTPPGGTGFAGHNSAICASSASY- 59

```

```

QY 60 INFEQPNPNLYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 1033 LGGIGVLTDPNLYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1092
QY 120 LEEVLTAKTVTVAPEAIGRTVQGTWQTKVQLPAGTKYVAFRHFQCTDFWMLDDVEI 179
Db 1093 LEETITAGK-VRSFKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQCTDFWMLDDVEI 1150
QY 180 KANGKRADEFTEPSSSTHGEAPAEWTTIDADGGQGLCLSSGQDMLTAHGSGNVVVSF 239
Db 1151 KANGKRADEFTEPSSSTHGEAPAEWTTIDADGGQGLCLSSGQDMLTAHGSGNVVVSF 1210
QY 240 SWNGMALPNPNLYLSKDVTKATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTWVFEET 299
Db 1211 SWNGMALPNPNLYLSKDVTKATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTWVFEET 1270
QY 300 PNGINKGGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 359
Db 1271 PNGINKGGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 1330
QY 360 TMGGSPTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCVN 419
Db 1331 TMGGSPTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCVN 1390
QY 420 VTVDVPQFNPVQNLTGSAV--GQKVTILKWDAP 449
Db 1391 VTVDVPQFNPVQNLTGSAV--GQKVTILKWDAP 1422

RESULT 11
Q51817 PRELIMINARY; PRT; 1732 AA.
AC Q51817;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Porphyrapain.
GN PRTP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=W12;
RX MEDLINE=96213011; PubMed=8631659;
RA Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
RA Progluske-Fox A., Lantz M.S.;
RT "Analysis of the prtp gene encoding porphyrapain, a cysteine proteinase
RT of Porphyromonas gingivalis."
RL J. Bacteriol. 178:2734-2741(1996).
DR EMBL; U42210; AAB06565.1; -.
DR PIR; T30836; T30836.
DR MEROPS; C25.002; -.
DR GO; GO:0005224; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25_1.
DR Pfam; PF03785; Peptidase_C25_C_1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1732 AA; 187875 MW; 654271DDBF7BCAE4 CRC64;

```

Query Match 82.3%; Score 2011.5; DB 2; Length 1732;
 Best Local Similarity 84.7%; Pred. No. 9e-124;
 Matches 393; Conservative 19; Mismatches 41; Indels 9; Gaps 6;

```

QY 1 PNPNGTTTSESPENGIPASWKTIIDADGGNNWTTTPPGCTSPAGHNSAICASSASY- 59
Db 977 PNPNG-TTSESPENGIPASWKTIIDADGGHGWKPGNAPG---IAGTNSNGCVYSFSG 1032
QY 60 INFEQPNPNLYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 1033 LGGIGVLTDPNLYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1092
QY 120 LEEVLTAKTVTVAPEAIGRTVQGTWQTKVQLPAGTKYVAFRHFQCTDFWMLDDVEI 179
Db 1093 LEETITAGK-VRSFKAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQCTDFWMLDDVEI 1150
QY 180 KANGKRADEFTEPSSSTHGEAPAEWTTIDADGGQGLCLSSGQDMLTAHGSGNVVVSF 239
Db 1151 KANGKRADEFTEPSSSTHGEAPAEWTTIDADGGQGLCLSSGQDMLTAHGSGNVVVSF 1210
QY 240 SWNGMALPNPNLYLSKDVTKATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTWVFEET 299
Db 1211 SWNGMALPNPNLYLSKDVTKATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTWVFEET 1270
QY 300 PNGINKGGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 359
Db 1271 PNGINKGGARFGLSTADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 1330
QY 360 TMGGSPTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCVN 419
Db 1331 TMGGSPTPTDYYTVYRDGDKIKEGLTETTFEEDGVATGNHEVCVEKYTAGVSPKCVN 1390
QY 420 VTVDVPQFNPVQNLTGSAV--GQKVTILKWDAP 449
Db 1391 VTVDVPQFNPVQNLTGSAV--GQKVTILKWDAP 1422

```

RESULT 12

```

O52050 PRELIMINARY; PRT; 1732 AA.
AC O52050;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Lysine specific cysteine protease.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=W83;
RX MEDLINE=98298016; PubMed=9632563;
RA Lewis J.P., Macrina F.L.;
RT "IS185, an insertion sequence-like element associated with protease
RT genes in Porphyromonas gingivalis."
RL Infect. Immun. 66:3035-3042(1998).
DR EMBL; AF017059; AAC26523.1; -.
DR MEROPS; C25.002; -.
DR GO; GO:0005224; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25_1.
DR Pfam; PF03785; Peptidase_C25_C_1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW Protease.
SQ SEQUENCE 1732 AA; 187931 MW; B2337463D5CB5EAS CRC64;

```

Query Match 81.8%; Score 1997.5; DB 2; Length 1732;

Best Local Similarity 84.3%; Pred. No. 7.5e-123;
Matches 381; Conservative 19; Mismatches 43; Indels 9; Gaps 6;

QY 1 PNPNGTTLSSSFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAICASSASY- 59
DB 977 PNPNG-TTLSSSFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAICASSASY- 59
QY 60 INFEQONPDNYLTPPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANAL 119
DB 1033 LGGIGVLTDPNYLITPALDPNGGKLTFWVCAQDANYASEHYAVASSTGNDASNFANAL 1092
QY 120 LSEVLTAQVTVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDFFWINLDDVEI 179
DB 1093 LSEITIAKG-VRSPKAIRG-RIQGTWRQKTVLPAGTKYVAFRHFCTDFFWINLDDVEI 1150
QY 180 KANGKRADETFEFTSTHGEAFAEWTTIDADGGQGLCLSSGQDLWLTAGHGTNNVASF 239
DB 1151 KANGKRADETFEFTSTHGEAFAEWTTIDADGGQGLCLSSGQDLWLTAGHGTNNVASF 1210
QY 240 SWNGMALNDPNYLI SKDVTGATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDFTVVFBE 299
DB 1211 SWNGMALNDPNYLI SKDVTGATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDFTVVFBE 1270
QY 300 PNGINKGARFGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDNLVILDDIQF 359
DB 1271 PNGINKGARFGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDNLVILDDIQF 1330
QY 360 TWGGSPTDITYTVYRGTGKIKEGLTETTFEEDGATGNHEYCVVEKYTAGVSPKCVN 419
DB 1331 TWGGSPTDITYTVYRGTGKIKEGLTETTFEEDGATGNHEYCVVEKYTAGVSPKCVN 419
QY 420 VTVDVQVNPQNLGSAV--GQKVTLKWDAP 449
DB 1391 VTVDVQVNPQNLGSAV--GQKVTLKWDAP 449

RESULT 13
Q9F470 ID Q9F4J0 PRELIMINARY; PRT; 925 AA.
AC Q9F4J0;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative outer membrane protein Pg57.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT "Identification of vaccine candidates from genomic analysis of
RT Porphyromonas gingivalis";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007522; AAG24228.1;
DR InterPro; IPR003961; FN III.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 925 AA; 103632 MW; 5FF2198D6914DAE8 CRC64;

Query Match 28.0%; Score 683; DB 2; Length 925;
Best Local Similarity 30.2%; Pred. No. 1.3e-36;
Matches 176; Conservative 62; Mismatches 132; Indels 212; Gaps 22;

QY 10 LSESFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAICASSASYINF 62

DB 311 LYENFENGVPNGWLVITADGGFSW-----GHYLNAYDAFPNGHGCHSLSASYVPG 363
QY 63 EGPQPNPNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANALBEE 122
DB 364 IGPVTPDNVLTIPKVE--GAKRVKYVSTQDANWAAEHYAVWASTTGTAVGDFV-ILFEE 420
QY 123 VLTAKTVTVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDFFWINLDDV- 177
DB 421 TMTAKPT-----GAWYERTINLPKGTKYIARHYNCTDIYFLKLLDDITVFT 467
QY 178 -----BIKANG----- 183
DB 468 PASEPEPTVDFVSLIENKGRKLNYPNGYEPDKTDDKPLQLAGYNIYANGSLLVHIQ 527
QY 184 -----KRAFTET-----FESSTHGE 199
DB 528 DPTVLEYIDETVSSRDDQVEVEYCVTAAYNDNISQSVCDKLIYDSQSDIILEGFEAGS 587
QY 200 APAEWTTIDADGGQGLCLSSGQDLW-LTAGHGTNNVAVSFWNGM--ALNPDNYLISKD 256
DB 588 IPEGWLLIDADGDNVW-----DYPTMYGHDSEKCIASPSYLPMLGVLTFEDNYLVTPR 642
QY 257 VTGATKVKYIYAVNDG-PPGDHYAVMISKTGTNAGDFTVVFEEETPNGINKGARFGLSTE 315
DB 643 LEGAKLVKYVSAQDAVYSAEHYAVWSTTGTAVDFVLLFEET-----MTAK 690
QY 316 ADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDNLVILDDI- 357
DB 691 ANGA-----WYERTITLPAQTKYIARHYDCTDMFILLDDITVYRSTETPEVPTDFV 745
QY 358 -----QFTMGSSPTPD-----YTVTVRDTG-----KIKEGLTETTF 390
DB 746 SLIENKGRKLNYPNGYEPDKTDDKPLQLAGYNIYANGSLLVHIQDPTVLEYIDETYS 805
QY 391 EEDGATGNHEYCVVEKYTAGV-SPKEC--VNVTV--DPVQ 426
DB 806 SRDQGV--EMECYCVTAAYNDNISQSVCDKLIYTSIDNIQ 845

RESULT 14
Q9KIB3 ID Q9KIB3 PRELIMINARY; PRT; 312 AA.
AC Q9KIB3;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237555; AAF81413.1;
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match 23.0%; Score 561.5; DB 2; Length 312;
Best Local Similarity 32.8%; Pred. No. 3.2e-29;
Matches 136; Conservative 23; Mismatches 73; Indels 183; Gaps 8;

QY 10 LSESFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAICASSASYINFEGQNP 69

```
Db 4 LSESPESGIPALWKTIDADGGYNMHL-----TNFTG-QSGLCVSSASYIGVGALTPD 57
QY 70 NYLVTPELSLPGG--TLTFWCAODANYASBHYAVYASSTGNDASNEANALLEEVLTA 127
Db 58 NYLITPELKLPTDALVEIIVWCTQDLTPSHYAVYSSSTGNNADFNLLYEETLTAK 117
QY 128 TVVTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFGCTDFWLNDDVEIKANGK 187
Db 118 R-IQSPELIRGNRTQGVWYQKWLPNDTKYVAFRHFNSTDFWLNLDVSL----- 169
QY 188 FTETFESTHGEPAPAEWITIDADGGQGWLCSSQGLDHLTAHGNTVNVASFWSWNGMA 247
Db 170 ----- 169
QY 248 PDNYLISKDVTGATKVYKYAVNDGPGDGHYAVMISKTGNAGDFTVVFEETPNGINKG 307
Db 170 -----YTP----- 172
QY 308 ARFGLSTEADGAKPOSVMIVRTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGSP 367
Db 173 -----LRAPC-----PHP 182
QY 368 TDYTVVVRDGTKEIGLTTFTFEEDGVATGNHGYCVVEKYTAGVSPKCVNVTV 422
Db 183 GGYTVSVFRDQKIASGLSALAYIDTDPYGTQYCVQVNYLQGDYKVKCKNIIV 237
```

RESULT 15

```
Q9XBU9 PRELIMINARY; PRT; 293 AA.
AC Q9XBU9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Immunoreactive 32 kDa antigen PG25.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
CX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr J., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF155351; AAD38410.1; -.
SQ SEQUENCE 293 AA; 32272 MW; CC03EAC241F7F6F1 CRC64;
```

```
Query Match 9.5%; Score 233; DB 2; Length 293;
Best Local Similarity 59.0%; Pred. No. 1.3e-07;
Matches 46; Conservative 7; Mismatches 23; Indels 2; Gaps 1;

QY 370 YTVTVYRDGKTKIKGLTETTFEEDGVATGNHGYCVVEKYTAGVSPKCVNVTVDPQFNP 429
Db 77 YTVTVYRDGKTKIKGLTETTFEEDGVATGNHGYCVVEKYTAGVSPKCVNVTVDPQFNP 429
QY 430 VQNLTSAGVQKVKLWD 447
Db 135 VTNLGTASNDVSLDWD 152
```

Search completed: May 18, 2004, 11:46:36
Job time : 34.8962 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 13.9034 Seconds
(without alignments)
1670.936 Million cell updates/sec

Title: US-08-570-311-16

Perfect score: 2443

Sequence: 1 PNNPGTTLSESPENGIIPA.....QLTGSAGVQKVLKWDAPN 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTGUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------------|
| 1 | 2443 | 100.0 | 450 | 2 | US-08-570-311-16 |
| 2 | 2443 | 100.0 | 2628 | 2 | US-08-570-311-14 |
| 3 | 2436 | 99.7 | 456 | 2 | US-08-570-311-18 |
| 4 | 2436 | 99.7 | 456 | 2 | US-08-570-311-20 |
| 5 | 2321 | 95.0 | 439 | 2 | US-08-570-311-22 |
| 6 | 2063 | 84.4 | 1477 | 4 | US-09-482-500A-1 |
| 7 | 2063 | 84.4 | 1687 | 2 | US-08-570-311-29 |
| 8 | 2063 | 84.4 | 1704 | 3 | US-08-336-308A-10 |
| 9 | 2063 | 84.4 | 1704 | 3 | US-08-822-324-6 |
| 10 | 2051 | 84.0 | 1704 | 3 | US-09-490-931-10 |
| 11 | 2051 | 84.0 | 1706 | 4 | US-09-066-330-10 |
| 12 | 2036 | 83.3 | 1087 | 2 | US-08-570-311-8 |
| 13 | 2036 | 83.3 | 1087 | 2 | US-08-353-485-8 |
| 14 | 2036 | 83.3 | 1358 | 2 | US-08-570-311-27 |
| 15 | 2016.5 | 82.5 | 1732 | 4 | US-09-066-330-11 |
| 16 | 2011.5 | 82.3 | 1732 | 2 | US-08-570-311-10 |
| 17 | 2011.5 | 82.3 | 1732 | 2 | US-08-353-485-10 |
| 18 | 546.5 | 22.4 | 497 | 2 | US-08-570-311-2 |
| 19 | 546.5 | 22.4 | 497 | 2 | US-08-353-485-2 |
| 20 | 296.5 | 12.1 | 942 | 1 | US-08-141-324-14 |
| 21 | 296.5 | 12.1 | 942 | 1 | US-08-541-902-14 |
| 22 | 236 | 9.7 | 49 | 3 | US-08-822-324-18 |
| 23 | 189 | 7.7 | 46 | 3 | US-08-822-324-9 |
| 24 | 135 | 5.5 | 1833 | 4 | US-08-621-944A-4 |
| 25 | 135 | 5.5 | 1833 | 4 | US-08-621-944A-4 |
| 26 | 135 | 5.5 | 1992 | 4 | US-08-945-567D-4 |
| 27 | 135 | 5.5 | 1992 | 4 | US-08-945-567D-3 |

28 133.5 5.5 2123 3 US-08-968-685A-10 Sequence 10, Appli
29 133.5 5.5 2736 4 US-09-252-991A-30227 Sequence 30227, A
30 130.5 5.3 492 4 US-09-482-500A-2 Sequence 2, Appli
31 130.5 5.3 737 1 US-08-119-361-5 Sequence 5, Appli
32 130.5 5.3 737 3 US-08-336-308A-4 Sequence 4, Appli
33 130.5 5.3 737 3 US-08-822-324-4 Sequence 4, Appli
34 130.5 5.3 737 3 US-09-490-931-4 Sequence 4, Appli
35 129 5.3 24 4 US-09-066-330-6 Sequence 6, Appli
36 128 5.2 25 3 US-08-822-324-19 Sequence 19, Appli
37 128 5.2 2315 4 US-08-543-681A-5434 Sequence 5434, Ap
38 125.5 5.1 1095 4 US-08-851-567B-34 Sequence 34, Appli
39 124.5 5.1 2411 4 US-09-268-347-36 Sequence 36, Appli
40 122 5.0 509 3 US-08-822-324-8 Sequence 8, Appli
41 121.5 5.0 1638 4 US-09-071-035-258 Sequence 258, App
42 121.5 5.0 1638 4 US-09-071-035-262 Sequence 262, App
43 121.5 5.0 1638 4 US-09-071-035-266 Sequence 266, App
44 121.5 5.0 1747 4 US-09-134-000C-5999 Sequence 5999, Ap
45 121.5 5.0 2048 4 US-09-268-347-48 Sequence 48, Appli

ALIGNMENTS

RESULT 1

US-08-570-311-16
; Sequence 16, Application US/08570311
; Patent No. 5824791

GENERAL INFORMATION:

APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph

TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ted W. Whitlock

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/570,311

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/353,485

FILING DATE: 09-DEC-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/647,119

FILING DATE: 25-JAN-1991

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/241,640

FILING DATE: 08-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UFL5.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 450 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-16

Query Match 100.0%; Score 2443; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 3.1e-203;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNPNGTTLSEFENGIPASWKTIDADGDNNTTTPPGTSPAGHNSAICASSAYI 60
Db 1 PNPNGTTLSEFENGIPASWKTIDADGDNNTTTPPGTSPAGHNSAICASSAYI 60

QY 61 NFEQPNPDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANALL 120
Db 61 NFEQPNPDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANALL 120

QY 121 EVLTKTAVTVAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWNLDDVEIK 180
Db 121 EVLTKTAVTVAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWNLDDVEIK 180

QY 181 ANGKRADETFETFSSTHGEAPAEWTTIDADGGQWCLSSGQDLWLTAGHTNVVASFS 240
Db 181 ANGKRADETFETFSSTHGEAPAEWTTIDADGGQWCLSSGQDLWLTAGHTNVVASFS 240

QY 241 WNGMALNPNDYLLISKDVGTATKVKYYAVNDGFGDHYAVMISKTGTNAGDFTVVFETP 300
Db 241 WNGMALNPNDYLLISKDVGTATKVKYYAVNDGFGDHYAVMISKTGTNAGDFTVVFETP 300

QY 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFFWNLDDIQT 360
Db 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFFWNLDDIQT 360

QY 361 MGSPTPTDYTYVYRDGTIKIEGLTETTFEDGATGNHCVKVTAGVSPKCVNV 420
Db 361 MGSPTPTDYTYVYRDGTIKIEGLTETTFEDGATGNHCVKVTAGVSPKCVNV 420

QY 421 TVDPVQFNPVQNLTSAGVQKVTWKWDAPN 450
Db 421 TVDPVQFNPVQNLTSAGVQKVTWKWDAPN 450

RESULT 2

US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-14

Query Match 100.0%; Score 2443; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 4.5e-202;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PNPNGTTLSEFENGIPASWKTIDADGDNNTTTPPGTSPAGHNSAICASSAYI 60
Db 500 PNPNGTTLSEFENGIPASWKTIDADGDNNTTTPPGTSPAGHNSAICASSAYI 559

QY 61 NFEQPNPDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANALL 120
Db 560 NFEQPNPDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANALL 619

QY 121 EVLTKTAVTVAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWNLDDVEIK 180
Db 620 EVLTKTAVTVAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWNLDDVEIK 679

QY 181 ANGKRADETFETFSSTHGEAPAEWTTIDADGGQWCLSSGQDLWLTAGHTNVVASFS 240
Db 680 ANGKRADETFETFSSTHGEAPAEWTTIDADGGQWCLSSGQDLWLTAGHTNVVASFS 739

QY 241 WNGMALNPNDYLLISKDVGTATKVKYYAVNDGFGDHYAVMISKTGTNAGDFTVVFETP 300
Db 740 WNGMALNPNDYLLISKDVGTATKVKYYAVNDGFGDHYAVMISKTGTNAGDFTVVFETP 799

QY 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFFWNLDDIQT 360
Db 800 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHHGCTDFFWNLDDIQT 859

QY 361 MGSPTPTDYTYVYRDGTIKIEGLTETTFEDGATGNHCVKVTAGVSPKCVNV 420
Db 860 MGSPTPTDYTYVYRDGTIKIEGLTETTFEDGATGNHCVKVTAGVSPKCVNV 919

QY 421 TVDPVQFNPVQNLTSAGVQKVTWKWDAPN 450
Db 920 TVDPVQFNPVQNLTSAGVQKVTWKWDAPN 949

RESULT 3
US-08-570-311-18
; Sequence 18, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn

QY 61 NFEQPNDNYLVTPELSLPGNGTLTFWVCAQDANYASHEHYAVYASSTGNDASNFANALL 120
DB 67 NFEQPNDNYLVTPELSLPGNGTLTFWVCAQDANYASHEHYAVYASSTGNDASNFANALL 126
QY 121 EEVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
DB 127 EEVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
QY 181 ANGKRADETFETPSSHGEPAEWTTIDADGQGWCLSSGQDNLTAHGGTNVVASFS 240
DB 187 ANGKRADETFETPSSHGEPAEWTTIDADGQGWCLSSGQDNLTAHGGTNVVASFS 246
QY 241 WNGMALNPNDYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEETP 300
DB 247 WNGMALNPNDYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEETP 306
QY 301 NGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 360
DB 307 NGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 366
QY 361 MGSPPTDITYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVEVKYTAGVSPKVCNV 420
DB 367 MGSPPTDITYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVEVKYTAGVSPKVCNV 426
QY 421 TVDPVQFNPVQNLTSAGVQKVTWKWDAPN 450
DB 427 TVDPVQFNPVQNLTSAGVQKVTWKWDAPN 456

RESULT 5

US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 439 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-22

Query Match 95.0%; Score 2321; DB 2; Length 439;
Best Local Similarity 98.4%; Pred. No. 1.1e-192;

Matches 426; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 PNPNGTITLSEFNGIPASWKTIDADGDNNTTTPPGGTSPAGHNSAICVSSASYI 60
DB 7 PNPNGTITLSEFNGIPASWKTIDADGDNNTTTPPGGTSPAGHNSAICVSSASYI 66
QY 61 NFEQPNDNYLVTPELSLPGNGTLTFWVCAQDANYASHEHYAVYASSTGNDASNFANALL 120
DB 67 NFEQPNDNYLVTPELSLPGNGTLTFWVCAQDANYASHEHYAVYASSTGNDASNFANALL 126
QY 121 EEVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 180
DB 127 EEVLTAKTAVTAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDDVEIK 186
QY 181 ANGKRADETFETPSSHGEPAEWTTIDADGQGWCLSSGQDNLTAHGGTNVVASFS 240
DB 187 ANGKRADETFETPSSHGEPAEWTTIDADGQGWCLSSGQDNLTAHGGTNVVASFS 246
QY 241 WNGMALNPNDYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEETP 300
DB 247 WNGMALNPNDYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEETP 306
QY 301 NGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 360
DB 307 NGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOFT 366
QY 361 MGSPPTDITYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVEVKYTAGVSPKVCNV 420
DB 367 MGSPPTDITYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVEVKYTAGVSPKVCNV 426
QY 421 TVDPVQFNPVQNL 433
DB 427 TINPTQFNPVQNL 439

RESULT 6

US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-09-482-500A-1

Query Match 84.4%; Score 2063; DB 4; Length 1477;

```
Best Local Similarity 86.3%; Pred. No. 1.5e-169;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGPTTLLSESPENGIPASWKTIIDADGGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Db 729 PNPNGPTTLLSESPENGIPASWKTIIDADGGHGWKPGNAPG---TAGYNSGCVYSESGF 785
QY 60 INFEGPQPNLYLTPVPELSPLNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 786 LGGIGVLTDPNLYLTPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 845
QY 120 LEEVLTAKTAVTAPRAIGTRVQGTWQKTVQVLPAGTKYVAFRHPGCTDFWNLDDVEI 179
Db 846 LEEVITAKG-VRSPEAIRG-RIQGTWQKTVQVLPAGTKYVAFRHPGCTDFWNLDDVEI 903
QY 180 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGTNVASF 239
Db 904 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGTNVASF 963
QY 240 SWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGNTAGDFTVVFEE 299
Db 964 SWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGNTAGDFTVVFEE 1023
QY 300 PNGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLYLLDDIQF 359
Db 1024 PNGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLYLLDDIQF 1083
QY 360 TMGGSPTPTDYTYVYRDGKIKGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKECVN 419
Db 1084 TMGGSPTPTDYTYVYRDGKIKGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKECVN 1143
QY 420 VTVDVPQFNPVQNLTGSAVGQKVLTKWDAPN 450
Db 1144 VTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1174

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulsk-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
```

```
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1687 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-29

Query Match 84.4%; Score 2063; DB 2; Length 1687;
Best Local Similarity 86.3%; Pred. No. 1.8e-169;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

QY 1 PNPNGPTTLLSESPENGIPASWKTIIDADGGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Db 939 PNPNGPTTLLSESPENGIPASWKTIIDADGGHGWKPGNAPG---TAGYNSGCVYSESGF 995
QY 60 INFEGPQPNLYLTPVPELSPLNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 996 LGGIGVLTDPNLYLTPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1055
QY 120 LEEVLTAKTAVTAPRAIGTRVQGTWQKTVQVLPAGTKYVAFRHPGCTDFWNLDDVEI 179
Db 1056 LEEVITAKG-VRSPEAIRG-RIQGTWQKTVQVLPAGTKYVAFRHPGCTDFWNLDDVEI 1113
QY 180 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGTNVASF 239
Db 1114 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGTNVASF 1173
QY 240 SWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGNTAGDFTVVFEE 299
Db 1174 SWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGNTAGDFTVVFEE 1233
QY 300 PNGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLYLLDDIQF 359
Db 1234 PNGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLYLLDDIQF 1293
QY 360 TMGGSPTPTDYTYVYRDGKIKGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKECVN 419
Db 1294 TMGGSPTPTDYTYVYRDGKIKGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKECVN 1353
QY 420 VTVDVPQFNPVQNLTGSAVGQKVLTKWDAPN 450
Db 1354 VTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1384

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
```

```

/ COUNTRY: US
/ ZIP: 80303
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/336,308A
/ FILING DATE: 08-NOV-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/119,361
/ FILING DATE: 10-SEP-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/265,441
/ FILING DATE: 24-JUN-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ferber, Donna M.
/ REGISTRATION NUMBER: 33,878
/ REFERENCE/DOCKET NUMBER: 21-93C
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 499-8080
/ TELEFAX: (303) 499-8089
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1704 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-336-308A-10

Query Match 84.4%; Score 2063; DB 3; Length 1704;
Best Local Similarity 86.3%; Pred. No. 1.9e-169;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

Qy 1 PNPNGTTLSESPENGIPASWKTIADGCGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Db 956 PNPNGTTLSESPENGIPASWKTIADGCGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Qy 60 INFEGPQPNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANAL 119
Db 1013 LGGIGVLTDPNLYLTALDLPNGGKLTFWVCAQDANYASEHYAVYASTGNDASNFANAL 1072
Qy 120 LEEVLTAKTAVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFQGTDFFWINLDDVEI 179
Db 1073 LEETITAGK-VRSPEAIRG-RIQGTWRQKTVQLPAGTKYVAFRHFQGTDFFWINLDDVEI 1130
Qy 180 KANGKRADFTETFESSTHGEAPAEWTTIDADGGQGWCLSSGQLDWLTAGGNTNVASF 239
Db 1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGGQGWCLSSGQLDWLTAGGNTNVASF 1190
Qy 240 SWNGMALNPONLYLSKDVGTGATKVYVAVNDGPPGDHYAVMISKTGNTAGDFTVPEET 299
Db 1191 SWNGMALNPONLYLSKDVGTGATKVYVAVNDGPPGDHYAVMISKTGNTAGDFTVPEET 1250
Qy 300 PNGINKGARFGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDLYILLDDIQF 359
Db 1251 PNGINKGARFGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDLYILLDDIQF 1310
Qy 360 TMGGSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECVN 419
Db 1311 TMGGSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECVN 1370
Qy 420 VTVPQPNVQNLGTGSAVGQVTLKWDAPN 450
Db 1371 VTINPTQPNVKNLKAQPDGDDVVLKNEAPS 1401

RESULT 9
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917

```

```

/ GENERAL INFORMATION:
/ APPLICANT: Potempa, Jan S.
/ APPLICANT: Travis, James
/ APPLICANT: Genco, Caroline A.
/ TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
/ TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
/ TITLE OF INVENTION: METHODS
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
/ STREET: 5370 Manhattan Circle, Suite 201
/ CITY: Boulder
/ STATE: CO
/ COUNTRY: US
/ ZIP: 80303
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/822,324
/ FILING DATE: 21-MAR-1997
/ CLASSIFICATION: 536
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/013,945
/ FILING DATE: 22-MAR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ferber, Donna M.
/ REGISTRATION NUMBER: 33,878
/ REFERENCE/DOCKET NUMBER: 103-95 WO
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 488-8080
/ TELEFAX: (303) 499-8089
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1704 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-822-324-6

Query Match 84.4%; Score 2063; DB 3; Length 1704;
Best Local Similarity 86.3%; Pred. No. 1.9e-169;
Matches 389; Conservative 17; Mismatches 39; Indels 6; Gaps 4;

Qy 1 PNPNGTTLSESPENGIPASWKTIADGCGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Db 956 PNPNGTTLSESPENGIPASWKTIADGCGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Qy 60 INFEGPQPNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASTGNDASNFANAL 119
Db 1013 LGGIGVLTDPNLYLTALDLPNGGKLTFWVCAQDANYASEHYAVYASTGNDASNFANAL 1072
Qy 120 LEEVLTAKTAVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFQGTDFFWINLDDVEI 179
Db 1073 LEETITAGK-VRSPEAIRG-RIQGTWRQKTVQLPAGTKYVAFRHFQGTDFFWINLDDVEI 1130
Qy 180 KANGKRADFTETFESSTHGEAPAEWTTIDADGGQGWCLSSGQLDWLTAGGNTNVASF 239
Db 1131 KANGKRADFTETFESSTHGEAPAEWTTIDADGGQGWCLSSGQLDWLTAGGNTNVASF 1190
Qy 240 SWNGMALNPONLYLSKDVGTGATKVYVAVNDGPPGDHYAVMISKTGNTAGDFTVPEET 299
Db 1191 SWNGMALNPONLYLSKDVGTGATKVYVAVNDGPPGDHYAVMISKTGNTAGDFTVPEET 1250
Qy 300 PNGINKGARFGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDLYILLDDIQF 359
Db 1251 PNGINKGARFGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNCSDLYILLDDIQF 1310
Qy 360 TMGGSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECVN 419
Db 1311 TMGGSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKECVN 1370

```


Db 1313 TMGSPPTDYTYVYRDGKIKKGLTETTFEEDGVAATGNHEVCVEVKYTAGVSPKVCVN 1372

QY 420 VTVDVQFNPVNLTGSAGVQKVTWKWDAPN 450

Db 1373 VTVNSTQFNPVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 12

US-08-570-311-8

Sequence 8, Application US/08570311

Patent No. 5824791

GENERAL INFORMATION:

APPLICANT: Progulsk-Fox, Ann

APPLICANT: Tumwasorn, Somying

APPLICANT: Lepine, Guylaine

APPLICANT: Han, Naiming

APPLICANT: Lantz, Marilyn

APPLICANT: Patti, Joseph

TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESS: Ted W. Whitlock

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/570,311

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/353,485

FILING DATE: 09-DEC-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/647,119

FILING DATE: 25-JAN-1991

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/241,640

FILING DATE: 08-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UF15.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1087 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-570-311-8

Query Match 83.3%; Score 2036; DB 2; Length 1087;

Best Local Similarity 85.4%; Pred. No. 2e-167;

Matches 385; Conservative 18; Mismatches 42; Indels 6; Gaps 4;

QY 1 PNPNGTTLSEFNGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASSASY- 59

Db 339 PNPNGTTLSEFNGIPASWKTIDADGCHGKGNAGF---IAGYNSGCVYSESG 395

QY 60 INFGQNPDPNYLVTPELSLPNGGTLTFWVCAQDANYASBEHYAVYASSTGNDASNFANAL 119

Db 396 LGGIGVLTDPNYLITPALDLANGKLTFFWVCAQDANYASBEHYAVYASSTGNDASNFANAL 455

QY 120 LEEVLAKTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFWNLDDVEI 179

Db 456 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPLPAGTKYVAFRHFQSTDMFYIDLDEVEI 513

QY 180 KANGKRADEFTETPESSTHGEAPAEWTTIDADGGQGWLCISSLGQDLWLTAGGTNNVVASF 239

Db 514 KANGKRADEFTETPESSTHGEAPAEWTTIDADGGQGWLCISSLGQDLWLTAGGTNNVVASF 573

QY 240 SWNGMALNPDPNLTISKDVTKATKYKYVAVNDGPPGPDHYAVMISKKTGNNAGDFTVVEET 299

Db 574 SWNGMALNPDPNLTISKDVTKATKYKYVAVNDGPPGPDHYAVMISKKTGNNAGDFTVVEET 633

QY 300 PNGINKGARGGLSTEADGAKPQSVWLTERTVDLPAGTKYVAFRHYNCSDNYILLDDIQF 359

Db 634 PNGINKGARGGLSTEADGAKPQSVWLTERTVDLPAGTKYVAFRHYNCSDNYILLDDIQF 693

QY 360 TMGSPPTDYTYVYRDGKIKKGLTETTFEEDGVAATGNHEVCVEVKYTAGVSPKVCVN 419

Db 694 TMGSPPTDYTYVYRDGKIKKGLTETTFEEDGVAATGNHEVCVEVKYTAGVSPKVCVN 753

QY 420 VTVDVQFNPVNLTGSAGVQKVTWKWDAPN 450

Db 754 VTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 784

RESULT 13

US-08-353-485-8

Sequence 8, Application US/08353485

Patent No. 5830710

GENERAL INFORMATION:

APPLICANT: Progulsk-Fox, Ann

APPLICANT: Tumwasorn, Somying

APPLICANT: Lepine, Guylaine

APPLICANT: Han, Naiming

APPLICANT: Lantz, Marilyn

APPLICANT: Patti, Joseph

TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESS: Ted W. Whitlock

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/353,485

FILING DATE: 09-DEC-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/647,119

FILING DATE: 25-JAN-1991

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/241,640

FILING DATE: 08-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UF15.C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1087 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-353-485-B

Query Match      83.3%; Score 2036; DB 2; Length 1087;
Best Local Similarity 85.4%; Pred. No. 2e-167;
Matches 385; Conservative 18; Mismatches 42; Indels 6; Gaps 4;

QY 1 PNPNGPTTLLSESPENGIPASWKTIDADGGNNWTTTPPGTSGFAGHNSAICASSASY- 59
Db 339 PNPNGPTTLLSESPENGIPASWKTIDADGGHGWKPGNAPG---IAGYNSNGCVYSESGF 395
QY 60 INFEGPQPNLYITPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 396 LGGIGVLTDPNLYITPALDLANGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 455
QY 120 LEEVLTAKTIVTAPAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFPGCTDFFWINLDDVEI 179
Db 456 LEEITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEI 513
QY 180 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQGLWCLSSGQDLWLTAGHGTNNVASF 239
Db 514 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQGLWCLSSGQDLWLTAGHGTNNVASF 573
QY 240 SWNGMALNPDPNLYISKDVTGATKYKYVAVNDGPGFDHYAVMISKTGTNAGDFTVVFEE 299
Db 574 SWNGMALNPDPNLYISKDVTGATKYKYVAVNDGPGFDHYAVMISKTGTNAGDFTVVFEE 633
QY 300 PNGINKGARGFLSTADGAKPOSVWIERVTLDPAGTKYVAFRHYNCSDLNYILLDDIOF 359
Db 634 PNGINKGARGFLSTADGAKPOSVWIERVTLDPAGTKYVAFRHYNCSDLNYILLDDIOF 693
QY 360 TMGGSPTPTDYYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVYKVTAGVSPKCVN 419
Db 694 TMGGSPTPTDYYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVYKVTAGVSPKCVN 753
QY 420 VTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 450
Db 754 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 784

RESULT 14
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progleke-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guyllaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1358 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-27

```

```

Query Match      83.3%; Score 2036; DB 2; Length 1358;
Best Local Similarity 85.4%; Pred. No. 2.9e-167;
Matches 385; Conservative 18; Mismatches 42; Indels 6; Gaps 4;

QY 1 PNPNGPTTLLSESPENGIPASWKTIDADGGNNWTTTPPGTSGFAGHNSAICASSASY- 59
Db 610 PNPNGPTTLLSESPENGIPASWKTIDADGGHGWKPGNAPG---IAGYNSNGCVYSESGF 666
QY 60 INFEGPQPNLYITPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 667 LGGIGVLTDPNLYITPALDLANGKLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 726
QY 120 LEEVLTAKTIVTAPAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFPGCTDFFWINLDDVEI 179
Db 727 LEEITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLDEVEI 784
QY 180 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQGLWCLSSGQDLWLTAGHGTNNVASF 239
Db 785 KANGKRADFTETPESSTHGEAPAEWTTIDADGGQGLWCLSSGQDLWLTAGHGTNNVASF 844
QY 240 SWNGMALNPDPNLYISKDVTGATKYKYVAVNDGPGFDHYAVMISKTGTNAGDFTVVFEE 299
Db 845 SWNGMALNPDPNLYISKDVTGATKYKYVAVNDGPGFDHYAVMISKTGTNAGDFTVVFEE 904
QY 300 PNGINKGARGFLSTADGAKPOSVWIERVTLDPAGTKYVAFRHYNCSDLNYILLDDIOF 359
Db 905 PNGINKGARGFLSTADGAKPOSVWIERVTLDPAGTKYVAFRHYNCSDLNYILLDDIOF 964
QY 360 TMGGSPTPTDYYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVYKVTAGVSPKCVN 419
Db 965 TMGGSPTPTDYYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVYKVTAGVSPKCVN 1024
QY 420 VTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 450
Db 1025 VTINPTQFNPVKNLKAQPDGGDVVLKWEAPS 1055

```

```

RESULT 15
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds

```

```
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match      82.5%; Score 2016.5; DB 4; Length 1732;
Best Local Similarity 85.0%; Pred. No. 2e-165;
Matches 384; Conservative 18; Mismatches 41; Indels 9; Gaps 6;

QY 1 PNPNCCTTTLSSEFFENGIPASMKTIIDADGDNWTTTTPPPGTSFAGHNSAICASSASY- 59
Db 977 PNPNG--TTLSSEFFENGIPASMKTIIDADGKGWKEGKNAPG---IAGYNSGCVISESG 1032

QY 60 INFEGPQNDNYLVTPELSIPLNGGTLTFWVCAQDANYASEHYAVYASSTGNDASFPANAL 119
Db 1033 LGGIGVLTDPDNYLITPDLPLNGGKLTFWVCAQDANYASEHYAVYASSTGNDASFTNAL 1092

QY 120 LEEVLTAKTVTAPAIIRGTRVGQWYQKTVQLPAGTKYVAFRHFQCTDFFWNLDDVEI 179
Db 1093 LEEITAKG-VRSFKAIRG-RIQGTWRQKTVDLPLAGTKYVAFRHFQSTDMFYIDLDEVEI 1150

QY 180 KANGKRADETFETPESSTHGEAPAEWTTIDADGGQGWCLSSGQLDWTAGHGTNVVASG 239
Db 1151 KANGKRADETFETPESSTHGEAPAEWTTIDADGGQGWCLSSGQLDWTAGHGSNVVSSF 1210

QY 240 SWNGMALNPDNYLISKDVTGATKVYKYAVNDGFGPDGHYAVMISKGTGNAGDFTVVFEET 299
Db 1211 SWNGMALNPDNYLISKDVTGATKVYKYAVNDGFGPDGHYAVMISKGTGNAGDFTVVFEET 1270

QY 300 PNGINKGARFGLSTADGAKPOSVMIERVTLPLAGTKYVAFRHYNCSDINYLILDDIQF 359
Db 1271 PNGINKGARFGLSTANGAKPOSVMIERVTLPLAGTKYVAFRHYNCSDINYLILDDIQF 1330

QY 360 TMGGSPTPDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHGYCVKKTAGVSPKECVN 419
Db 1331 TMGGSPTPDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHGYCVKKTAGVSPKCCVN 1390

QY 420 VTVDVPQFNPVQNLGSAV--GOKVTLKNDAP 449
Db 1391 VTVNSTQFNPVQNLTAEQAPNSMDAILKKNAP 1422
```

Search completed: May 18, 2004, 11:49:13
Job time : 14.9034 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:42:50 ; Search time 34.8564 Seconds
(without alignments)

3592.387 Million cell updates/sec

Title: US-08-570-311-16

Perfect score: 2443

Sequence: 1 PNPNGTTLTSEFENGIPA.....QLTGSVAGQVKTLKWDAPN 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 2051 | 84.0 | 1706 | 14 | US-10-229-066-10 |
| 2 | 2016.5 | 82.5 | 1732 | 14 | US-10-229-066-11 |
| 3 | 700 | 28.7 | 419 | 15 | US-10-174-695-3 |
| 4 | 680.5 | 27.9 | 419 | 15 | US-10-174-695-5 |
| 5 | 385 | 15.8 | 231 | 15 | US-10-174-695-6 |
| 6 | 206.5 | 8.5 | 196 | 15 | US-10-174-695-4 |
| 7 | 153 | 6.3 | 29 | 15 | US-10-387-977-18 |
| 8 | 153 | 6.3 | 29 | 15 | US-10-387-977-21 |
| 9 | 153 | 6.3 | 29 | 15 | US-10-387-977-23 |
| 10 | 153 | 6.3 | 29 | 15 | US-10-387-977-24 |
| 11 | 153 | 6.3 | 29 | 15 | US-10-387-977-26 |
| 12 | 148 | 6.1 | 872 | 12 | US-10-282-122A-55467 |
| 13 | 147 | 6.0 | 29 | 15 | US-10-387-977-17 |
| 14 | 145 | 5.9 | 29 | 15 | US-10-387-977-25 |
| 15 | 144 | 5.9 | 509 | 15 | US-10-387-977-101 |

| | | | | | | |
|----|-------|-----|------|----|----------------------|-------------------|
| 16 | 136 | 5.6 | 29 | 15 | US-10-387-977-20 | Sequence 20, Appl |
| 17 | 135 | 5.5 | 1833 | 14 | US-10-175-282-4 | Sequence 4, Appl1 |
| 18 | 135 | 5.5 | 1833 | 14 | US-10-175-275-4 | Sequence 4, Appl1 |
| 19 | 135 | 5.5 | 1992 | 14 | US-10-175-282-3 | Sequence 3, Appl1 |
| 20 | 135 | 5.5 | 1992 | 14 | US-10-175-275-3 | Sequence 3, Appl1 |
| 21 | 134 | 5.5 | 25 | 15 | US-10-387-977-77 | Sequence 77, Appl |
| 22 | 134 | 5.5 | 25 | 15 | US-10-387-977-80 | Sequence 80, Appl |
| 23 | 134 | 5.5 | 27 | 15 | US-10-387-977-11 | Sequence 11, Appl |
| 24 | 134 | 5.5 | 27 | 15 | US-10-387-977-12 | Sequence 12, Appl |
| 25 | 134 | 5.5 | 27 | 15 | US-10-387-977-13 | Sequence 13, Appl |
| 26 | 133.5 | 5.5 | 2122 | 9 | US-09-813-214A-9 | Sequence 9, Appl1 |
| 27 | 130.5 | 5.3 | 2468 | 12 | US-10-282-122A-66335 | Sequence 66335, A |
| 28 | 130.5 | 5.3 | 2468 | 14 | US-10-246-330-4 | Sequence 4, Appl1 |
| 29 | 130 | 5.3 | 25 | 15 | US-10-387-977-81 | Sequence 81, Appl |
| 30 | 129 | 5.3 | 24 | 14 | US-10-229-066-6 | Sequence 6, Appl1 |
| 31 | 129 | 5.3 | 2435 | 12 | US-10-282-122A-47453 | Sequence 47453, A |
| 32 | 128 | 5.2 | 1946 | 12 | US-10-282-122A-62947 | Sequence 62947, A |
| 33 | 127 | 5.2 | 25 | 15 | US-10-387-977-83 | Sequence 83, Appl |
| 34 | 127 | 5.2 | 25 | 15 | US-10-387-977-84 | Sequence 84, Appl |
| 35 | 127 | 5.2 | 25 | 15 | US-10-387-977-85 | Sequence 85, Appl |
| 36 | 126 | 5.2 | 1649 | 15 | US-10-369-493-18460 | Sequence 18460, A |
| 37 | 125.5 | 5.1 | 555 | 14 | US-10-156-761-14592 | Sequence 14592, A |
| 38 | 125.5 | 5.1 | 1095 | 14 | US-10-242-056-34 | Sequence 34, Appl |
| 39 | 125.5 | 5.1 | 1095 | 15 | US-10-262-794A-34 | Sequence 34, Appl |
| 40 | 124.5 | 5.1 | 691 | 15 | US-10-369-493-23473 | Sequence 23473, A |
| 41 | 124.5 | 5.1 | 1333 | 12 | US-10-282-122A-55546 | Sequence 55546, A |
| 42 | 123 | 5.0 | 31 | 15 | US-10-387-977-6 | Sequence 6, Appl1 |
| 43 | 122 | 5.0 | 491 | 15 | US-10-387-977-100 | Sequence 100, App |
| 44 | 121.5 | 5.0 | 1638 | 12 | US-10-206-576-258 | Sequence 258, App |
| 45 | 121.5 | 5.0 | 1638 | 12 | US-10-206-576-262 | Sequence 262, App |

ALIGNMENTS

RESULT 1

US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 84.0%; Score 2051; DB 14; Length 1706;
Best Local Similarity 85.8%; Pred. No. 9.8e-176;
Matches 387; Conservative 18; Mismatches 40; Indels 6; Gaps 4;

| | | |
|----|------|---|
| QY | 1 | PNPNGTTLTSEFENGIPASWKTTDADGDGNNWTTTPPPGGTSPAGHNSATCASSAY- 59 |
| DB | 958 | PNPNGTTLTSEFENGIPASWKTTDADGDGNNWTTTPPPGGTSPAGHNSATCASSAY- 1014 |
| QY | 60 | INFEQPNDNLTIVLPESLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119 |
| DB | 1015 | IGGIGVLTFDNLTIVLPESLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1074 |

QY 120 LEEVLTAKTVTVAPEAIRTRVQGTWYKTVQVLPAGTKYVAFRHFQCTDFFWINLDDVEI 179
Db 1075 LEETITAKG-VRSPEAMRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFYIDLDEVEI 1132
QY 180 KANGKRADFTETFESSHGAPAEWTTIDADGQGMCLSSGQDMLTAHGGTNVVASF 239
Db 1133 KANGKRADFTETFESSHGAPAEWTTIDADGQGMCLSSGQDMLTAHGGTNVVSF 1192
QY 240 SWNGMALNPONLYISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEE 299
Db 1193 SWNGMALNPONLYISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEE 1252
QY 300 PNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 359
Db 1253 PNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 1312
QY 360 TMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKBCVN 419
Db 1313 TMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKBCVN 1372
QY 420 VTVDPOFNPVQNLTSAGVQKVTILKWDAPN 450
Db 1373 VTVNSTQFNPVQNLKQAQPDGDDVVLKWEAPS 1403
RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Shakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11
Query Match 82.5%; Score 2016.5; DB 14; Length 1732;
Best Local Similarity 85.0%; Pred. No. 1.3e-172;
Matches 384; Conservative 18; Mismatches 41; Indels 9; Gaps 6;
QY 1 PNPNGTTLTSESFENGIPASWKTIDADGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Db 977 PNPNG-TTLESFENGIPASWKTIDADGNGKPGNAPG---IAGYNSGCVYSESFG 1032
QY 60 INFEGQNPNDNYLVTPELSLNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 1033 LGGIGVLTDPDNYLITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASNFANAL 1092
QY 120 LEEVLTAKTVTVAPEAIRTRVQGTWYKTVQVLPAGTKYVAFRHFQCTDFFWINLDDVEI 179
Db 1093 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFYIDLDEVEI 1150
QY 180 KANGKRADFTETFESSHGAPAEWTTIDADGQGMCLSSGQDMLTAHGGTNVVASF 239
Db 1151 KANGKRADFTETFESSHGAPAEWTTIDADGQGMCLSSGQDMLTAHGGTNVVSF 1210
QY 240 SWNGMALNPONLYISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEE 299

Db 1211 SWNGMALNPONLYISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVVFEE 1270
QY 300 PNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 359
Db 1271 PNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILLDDIOF 1330
QY 360 TMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKBCVN 419
Db 1331 TMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKBCVN 1390
QY 420 VTVDPOFNPVQNLTSAGV--GQKVTILKWDAP 449
Db 1391 VTVNSTQFNPVQNLTAQAPNSMDAILKWNAP 1422
RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Shakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3
Query Match 28.7%; Score 700; DB 15; Length 419;
Best Local Similarity 74.7%; Pred. No. 1.8e-54;
Matches 139; Conservative 11; Mismatches 30; Indels 6; Gaps 4;
QY 1 PNPNGTTLTSESFENGIPASWKTIDADGNNWTTTPPGGTSPAGHNSAICASSASY- 59
Db 239 PNPNGTTLTSESFENGIPASWKTIDADGNGKPGNAPG---IAGYNSGCVYSESFG 295
QY 60 INFEGQNPNDNYLVTPELSLNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANAL 119
Db 296 LGGIGVLTDPDNYLITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASNFANAL 355
QY 120 LEEVLTAKTVTVAPEAIRTRVQGTWYKTVQVLPAGTKYVAFRHFQCTDFFWINLDDVEI 179
Db 356 LEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFYIDLDEVEI 413
QY 180 KANGKR 185
Db 414 KANGKR 419
RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Shakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695

; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match 6.3%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKKGLTETTFEEEDGVAT 397
|||||
Db 1 DYTIVYRDGKIKKGLTETTFEEEDGVAT 29

RESULT 8

US-10-387-977-21
; Sequence 21, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match 6.3%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKKGLTETTFEEEDGVAT 397
|||||
Db 1 DYTIVYRDGKIKKGLTETTFEEEDGVAT 29

RESULT 9

US-10-387-977-23
; Sequence 23, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528

; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match 6.3%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKKGLTETTFEEEDGVAT 397
|||||
Db 1 DYTIVYRDGKIKKGLTETTFEEEDGVAT 29

RESULT 10

US-10-387-977-24
; Sequence 24, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match 6.3%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKKGLTETTFEEEDGVAT 397
|||||
Db 1 DYTIVYRDGKIKKGLTETTFEEEDGVAT 29

RESULT 11

US-10-387-977-26
; Sequence 26, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30

;; PRIOR APPLICATION NUMBER: AU PO 6528
;; PRIOR FILING DATE: 1997-04-30
;; NUMBER OF SEQ ID NOS: 105
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 26
;; LENGTH: 29
;; TYPE: PRT
;; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match 6.3%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKKGLTETTFEEDGVAT 397
Db 1 DYTIVYRDGKIKKGLTETTFEEDGVAT 29

RESULT 12

US-10-282-122A-55467
;; Sequence 55467, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangsu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.

;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 55467
;; LENGTH: 872
;; TYPE: PRT
;; ORGANISM: Enterobacter cloacae
US-10-282-122A-55467

Query Match 6.1%; Score 148; DB 12; Length 872;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 115; Conservative 59; Mismatches 196; Indels 160; Gaps 26;

QY 4 NPGTTTSSBENGIPASWKTTIDADGDNWTTTPPPGGTSFAGHNSAICASSASYINFE 63
Db 293 NPGTGHNVTVDTGLPS--VAFNAISDDNVLAWEKQDLSVSGTSANLAEGTVIVITLN 350
QY 64 GPONPDYLVTPBELSLPNGSTLTFWYCAQDANVASEHYAVASST---GNDASNFAALL 120
Db 351 G-----KNYAATTAAD---GTWSLTVPAAADLAGLDHYTLISATATNGVGSVNTANLLV 402
QY 121 EEVLTAKT-----VVTAPAIRGRVQGTWYQKTVQLPAG---TKYVAFRHFQGT-- 167
Db 403 DTALPTVTINTVAGDNVINAEEVAAGQTISG-----KVANAEGNTVTIIGNSYATVQ 458
QY 168 -DFFW-INLDDVEIKANGKRAADTFETFSSTHGEAPAEWTTIDADGGQWMLCLSSGOLD 225
Db 459 SDLTWSVNPESVLTALNGDLTVSATVINGHGTGAGEREIVIDAS-----LPGLRID 512
QY 226 WL-----TAHGGTVVASFSWNGMALPNPDNLYISKDVTGATKVXYVAVNDGFPDHY 278
Db 513 TVAGDDVINSIEHQNLIVTSSDGLA-----AGTTLT---VTVN---GKTY 553
QY 279 AVMISKTGT-----NAGDFTVVFEETPNGINKGARFGLSTEADGAKPOSVM 325
Db 554 AASVLADGTWSAIPAADVGALAACTTV---TAAGQSAG-----NPTVIS 597
QY 326 IERTVDLPAGTKYVAFRHYNCSDIANYLLDDI-----QFTMGGSPTPTD----- 369
Db 598 HDVTVDLSA-----VAI-----SIDAATDDVINAEEKGADLVLSGSTSVNEENQVTIT 647
QY 370 -----YTYTVYRDG-----TKIKEG-----LTEITFEEDGVATGNHHCVEVKYT 409
Db 648 SGGKTYTAKVDADGNWTATVPSADLAGLKDGDASQVSVTNAHNSASAGREYSVD---- 703
QY 410 AGVSPKECVNVTVDPVQFNPQNLTGSA-----VGOKVTLKWD 447
Db 704 -ATAP-----TVIDTVAGDNVINAESAAGVAISGTTAEVGTQVTVTLTD 748

RESULT 13

US-10-387-977-17
;; Sequence 17, Application US/10387977
;; Publication No. US20040005276A1
;; GENERAL INFORMATION:
;; APPLICANT: Reynolds, Eric Charles
;; APPLICANT: O'Brien-Simpson, Neil Martin
;; APPLICANT: Slakeski, Nada
;; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;; FILE REFERENCE: PORPHYROMONAS GINGIVALIS
;; CURRENT APPLICATION NUMBER: US/10/387,977
;; CURRENT FILING DATE: 2003-07-18
;; PRIOR APPLICATION NUMBER: US 09/423,056
;; PRIOR FILING DATE: 2000-03-22
;; PRIOR APPLICATION NUMBER: PCT/AU98/00311
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: AU PO 6528
;; PRIOR FILING DATE: 1997-04-30
;; NUMBER OF SEQ ID NOS: 105
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 17
;; LENGTH: 29
;; TYPE: PRT
;; ORGANISM: Porphyromonas gingivalis
US-10-387-977-17

Query Match 6.0%; Score 147; DB 15; Length 29;
Best Local Similarity 96.6%; Pred. No. 3.6e-06;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKKGLTETTFEEDGVAT 397
Db 1 DYTIVYRDGKIKKGLTETTFEEDGVAT 29

RESULT 14
 US-10-387-977-25
 ; Sequence 25, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 25
 ; LENGTH: 29
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-25

Query Match 5.9%; Score 145; DB 15; Length 29;
 Best Local Similarity 96.6%; Pred. No. 5.4e-06;
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 369 DYTIVYRDGKIKGLTETTFEEDGVAT 397
 Db 1 DYTIVYRDGKIKGLTETTFEEDGVAT 29

RESULT 15
 US-10-387-977-101
 ; Sequence 101, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 101
 ; LENGTH: 509
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-101

Query Match 5.9%; Score 144; DB 15; Length 509;
 Best Local Similarity 20.0%; Pred. No. 0.00043;
 Matches 94; Conservative 66; Mismatches 189; Indels 120; Gaps 21;

QY 48 HNSAICASSAS-VINFEQPNPNLVLPGLTTFWVCAODANYASHYAV-YA 105
 Db 67 YNDGLAASAPVFLAVG----DTDVISGEKGGKTKVTDLYISAVDGDYFPEMYFRMS 122

QY 106 SSTGNDASNEFANALL---EEVLTAKTVTVAPEAIRGTRVQGTWYQKTVQ--LPAGTKYVA 160
 Db 123 ASPEELTNLIIDKVLMEKATMPDKSYLEKVLIIAG--ADYSNNSOVGQPTIKYGMQYYY 180
 QY 161 FRHFGCTDFFWINLDDVEIKANGKRADFTTTPSSSTHGEAPAEWTTTIDADGGGWL--C 218
 Db 181 NQEHGYTDVY-----NYLKAPYTCYSHLNTGVSFANYT---AHGSETAWADPL 226
 QY 219 LSSQOLDWLT-----AHGGTNVVASPSNNGMALNPDNYLLSKDVTGATKVKYIYAVND 271
 Db 227 LTTSQLKALTNKKYFLAIGNCCITAFDY-----VQP-----CFGEVITRVKEKAGAYIIG 278
 QY 272 GFPGDH-----YAVMISKTGNAGDFTVVF--EETPNGINKGARGFLSTEA 316
 Db 279 SSPNSYWGEDIYWSVGANAVFGVQPTFEGTSMGSDATFLEDSYNTVN----- 326
 QY 317 DGAKPQSVWERTVDLPAQT-----KYVAFRHYNCSDNLNILLDDIOFTMGSPPTDYT- 371
 Db 327 -----SIMWAGNLAAATHAGNIGNITHIGAHYY--WEAYHVLGDSVMPYRAMPKNTYTL 379
 QY 372 -----YTVYRDGKIKGLTETTFEEDGVAT-----GNHE 401
 Db 380 PASLPQNASYSIOASAGSYVAISKOGVLYGTGVNAS-----GVATVSMTKQITENGNYD 435
 QY 402 YCDEVKYTAGVSPKECYNNVTVDPVQFNPVQNLITGSAYGQKVTLKWDAPN 450
 Db 436 VVITRSNYLPVIKIQVG---EPSPYQFVSNLTATTQGGKVTLKWEAPS 481

Search completed: May 18, 2004, 11:52:20
 Job time : 35.8564 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:33:39 ; Search time 47.0287 Seconds
(without alignments)
2739.638 Million cell updates/sec

Title: US-08-570-311-18

Perfect score: 2480

Sequence: 1 GTPNPNPNPNFTLSESF.....QNLTSAGVQKVTLLKWDAPN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|----------|--------------------|
| 1 | 2480 | 100.0 | 456 | 2 | AAR96023 | Aar96023 P. gingiv |
| 2 | 2480 | 100.0 | 456 | 2 | AAR96022 | Aar96022 P. gingiv |
| 3 | 2480 | 100.0 | 456 | 2 | AAW69491 | Aaw69491 Haemagglu |
| 4 | 2480 | 100.0 | 456 | 2 | AAW69490 | Aaw69490 Haemagglu |
| 5 | 2480 | 100.0 | 2628 | 2 | AAR96030 | Aar96030 P. gingiv |
| 6 | 2480 | 100.0 | 2628 | 2 | AAW69488 | Aaw69488 Haemagglu |
| 7 | 2436 | 98.2 | 450 | 2 | AAR96021 | Aar96021 P. gingiv |
| 8 | 2436 | 98.2 | 450 | 2 | AAW69489 | Aaw69489 Haemagglu |
| 9 | 2351 | 94.8 | 439 | 2 | AAR96024 | Aar96024 P. gingiv |
| 10 | 2351 | 94.8 | 439 | 2 | AAW69492 | Aaw69492 Haemagglu |
| 11 | 2082 | 84.0 | 1687 | 2 | AAR96033 | Aar96033 P. gingiv |
| 12 | 2082 | 84.0 | 1687 | 2 | AAW69495 | Aaw69495 Haemagglu |
| 13 | 2082 | 84.0 | 1704 | 2 | AAW69486 | Aaw69486 Haemagglu |
| 14 | 2082 | 84.0 | 1704 | 2 | AAW69483 | Aaw69483 Arg-gingi |
| 15 | 2082 | 84.0 | 1704 | 3 | AAW67396 | Aaw67396 Arg-gingi |
| 16 | 2082 | 84.0 | 1704 | 4 | AAU08938 | Aau08938 P. gingiv |
| 17 | 2070 | 83.5 | 1706 | 2 | AAW24786 | Aaw24786 PrtR anti |
| 18 | 2066 | 83.3 | 1087 | 2 | AAR96028 | Aar96028 P. gingiv |
| 19 | 2066 | 83.3 | 1087 | 2 | AAW69486 | Aaw69486 Haemagglu |
| 20 | 2066 | 83.3 | 1358 | 2 | AAR96032 | Aar96032 P. gingiv |
| 21 | 2066 | 83.3 | 1358 | 2 | AAW69494 | Aaw69494 Haemagglu |
| 22 | 2034.5 | 82.0 | 1732 | 2 | AAR96029 | Aar96029 P. gingiv |
| 23 | 2034.5 | 82.0 | 1732 | 2 | AAW24787 | Aaw24787 PrtK anti |
| 24 | 2034.5 | 82.0 | 1732 | 2 | AAW69487 | Aaw69487 Haemagglu |
| 25 | 726 | 29.3 | 419 | 4 | AAU03572 | Aau03572 P. gingiv |

ALIGNMENTS

RESULT 1

AAR96023
ID AAR96023 standard; protein; 456 AA.

XX AC AAR96023;

XX XX

DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

DE P. gingivalis haemagglutinin hAgA HAre3 product.

XX KW Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; HAre3.

XX OS Porphyromonas gingivalis; strain 381.

XX PN WO9617936-A2.

XX PD 13-JUN-1996.

XX PF 11-DEC-1995; 95WO-US016108.

XX PR 09-DEC-1994; 94US-00353485.

XX PA (UYFL) UNIV FLORIDA.

XX PA (UABR-) UAB RES FOUND.

XX PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX DR WPI; 1996-287181/29.

XX DR N-PSDB; AAT30647.

XX PT Porphyromonas gingivalis genes and proteins - used in the detection and vaccination against periodontal disease.

XX PS Claim 4; Page 110-112; 153pp; English.

XX CC HAre3 (AAR96023) is the product of the HAre3 repeat unit (AAT30647) of the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of haemagglutinin hAgA (see also AAR96030). HAre3 and other hAgA repeat unit products (see also AAR96021-22 and AAR96024) can be obtd. from transformed host cells and used as vaccines to protect humans or animals against periodontal disease. Expression in Salmonella cells allows produ. of live vaccine. HAre3-4 can also be used to detect the presence of anti-P. gingivalis antibodies and to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)

XX SQ Sequence 456 AA;

Abp55081 Porphyrom
Aau03574 P. gingiv
Aab49217 Peptide u
Aay34522 Porphorym
Aay34520 Porphorym
Aay34392 Porphorym
Aay34359 Porphorym
Aay34484 Porphorym
Aay36025 P. gingiv
Aaw69483 Haemagglu
Aar72458 Porphorym
Aau03575 P. gingiv
Aar77313 Porphorym
Aaw34805 Arg-speci
Aay34483 Porphorym
Aay34358 Porphorym
Aau03573 P. gingiv
Aaw34798 Arg-speci
Aaw83085 Peptide f

26 720 29.0 135 6 ABP55081
27 710.5 28.6 419 4 AAU03574
28 704 28.4 134 4 AAB49217
29 682 27.5 921 2 AAY34522
30 682 27.5 922 2 AAY34521
31 682 27.5 925 2 AAY34520
32 682 27.5 938 2 AAY34392
33 565.5 22.8 377 2 AAY34359
34 561.5 22.6 312 2 AAY34484
35 546.5 22.0 497 2 AAR96025
36 546.5 22.0 497 2 AAW69483
37 415 16.7 970 2 AAR72458
38 385 15.5 231 4 AAU03575
39 267 10.8 991 2 AAR77313
40 236 9.5 49 2 AAW34805
41 233 9.4 293 2 AAY34483
42 233 9.4 299 2 AAY34358
43 232.5 9.4 196 4 AAU03573
44 189 7.6 46 2 AAW34798
45 153 6.2 29 2 AAW83085

Query Match 100.0%; Score 2480; DB 2; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.5e-191;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTTILSESPENGI PASWKTIDADGGNNWTTTPPGGTSFAGHNSAICA 60
 Db 1 GTPNPNPNPGTTTILSESPENGI PASWKTIDADGGNNWTTTPPGGTSFAGHNSAICA 60

QY 61 SSASYINFEQPNPNLYVTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 Db 61 SSASYINFEQPNPNLYVTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

QY 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFCTDFFWINL 180
 Db 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFCTDFFWINL 180

QY 181 DDVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTN 240
 Db 181 DDVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTN 240

QY 241 VVASFWSNGMALPNPNLYLSKDVGTATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
 Db 241 VVASFWSNGMALPNPNLYLSKDVGTATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

QY 301 VPEETPNKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
 Db 301 VPEETPNKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

QY 361 DDQFTMGSPPTDITYTVYRDGTXI KEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
 Db 361 DDQFTMGSPPTDITYTVYRDGTXI KEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420

QY 421 KECUNVTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 456
 Db 421 KECUNVTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 456

RESULT 2
 AAR96022
 ID AAR96022 standard; protein; 456 AA.
 AC AAR96022;
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX P. gingivalis haemagglutinin hgaA Harep2 product.
 DE P. gingivalis haemagglutinin hgaA Harep2 product.
 KW Haemagglutinin; hgaA; periodontal disease; vaccine; antibody; Harep2.
 OS Porphyromonas gingivalis; strain 381.
 XX WO9617936-A2.
 XX 13-JUN-1996.
 XX 11-DEC-1995; 95WO-US016108.
 XX 09-DEC-1994; 94US-00353485.
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX Progukske-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX WPI; 1996-287181/29.
 DR N-PSDB; AAT30646.
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX Claim 4; Page 107-108; 153pp; English.
 PS
 XX

CC Harep2 (AAR96022) is the product of the Harep2 repeat unit (AAT30646) of
 CC the hgaA gene (AAT30654) of P. gingivalis 318. It forms part of
 CC haemagglutinin hgaA (see also AAR96030). Harep2 and other hgaA repeat
 CC unit products (see also AAR96021 and AAR96023-24) can be obtd. from
 CC transformed host cells and used as vaccines to protect humans or animals
 CC against periodontal disease. Expression in Salmonella cells allows prodn.
 CC of live vaccine. Harep1-4 can also be used to detect the presence of anti
 CC -P. gingivalis antibodies and to raise monoclonal antibodies for
 CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 456 AA;

Query Match 100.0%; Score 2480; DB 2; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.5e-191;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTTILSESPENGI PASWKTIDADGGNNWTTTPPGGTSFAGHNSAICA 60
 Db 1 GTPNPNPNPGTTTILSESPENGI PASWKTIDADGGNNWTTTPPGGTSFAGHNSAICA 60

QY 61 SSASYINFEQPNPNLYVTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 Db 61 SSASYINFEQPNPNLYVTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

QY 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFCTDFFWINL 180
 Db 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFCTDFFWINL 180

QY 181 DDVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTN 240
 Db 181 DDVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTN 240

QY 241 VVASFWSNGMALPNPNLYLSKDVGTATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300
 Db 241 VVASFWSNGMALPNPNLYLSKDVGTATKVKYYAVNDGFPGDHYAVMISKTGTNAGDFTV 300

QY 301 VPEETPNKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
 Db 301 VPEETPNKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

QY 361 DDQFTMGSPPTDITYTVYRDGTXI KEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
 Db 361 DDQFTMGSPPTDITYTVYRDGTXI KEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420

QY 421 KECUNVTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 456
 Db 421 KECUNVTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 456

RESULT 3
 AAW69491
 ID AAW69491 standard; protein; 456 AA.
 AC AAW69491;
 DT 22-DEC-1998 (first entry)
 DE Haemagglutinin protein hgaA, Harep3.
 KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 OS Porphyromonas gingivalis.
 XX US5824791-A.
 XX 20-OCT-1998.
 XX 11-DEC-1995; 95US-00570311.
 XX 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX

PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
 XX 20-OCT-1998.
 XX 11-DEC-1995; 95US-00570311.
 DR WPI; 1998-582627/49.
 DR N-PSDB; AAV58878.
 XX 08-SEP-1988; 88US-00241640.
 XX 25-JAN-1991; 91US-00647119.
 XX 09-DEC-1994; 94US-00353485.
 PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease poly.peptide(s)).
 XX (UYFL) UNIV FLORIDA.
 XX (UABR-) UAB RES FOUND.
 PS Claim 1; Col 133-138; 101pp; English.
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX Sequence 456 AA;
 SQ
 Query Match 100.0%; Score 2480; DB 2; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.5e-191;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTPNPNPNPGTITLSEFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICA 60
 DB 1 GTPNPNPNPGTITLSEFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICA 60
 QY 61 SSASYINPEGPONPDNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 DB 61 SSASYINPEGPONPDNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 QY 121 FANALLEEVLTAKTVVTAPPAIRGRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
 DB 121 FANALLEEVLTAKTVVTAPPAIRGRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
 QY 181 DDVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGGTN 240
 DB 181 DDVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGGTN 240
 QY 241 VVASFWSNGMALNPNDNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKGTNAGDFTV 300
 DB 241 VVASFWSNGMALNPNDNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKGTNAGDFTV 300
 QY 301 VFEETPNKNGGARFGLSTEADGAKPQSVWIERVTDLPACTKYVAFRHYNCSDLNYILL 360
 DB 301 VFEETPNKNGGARFGLSTEADGAKPQSVWIERVTDLPACTKYVAFRHYNCSDLNYILL 360
 QY 361 DDIOFTMGSSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 DB 361 DDIOFTMGSSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 QY 421 KECVNVTVDVPQFNPVQNLTGSAVGQKVTWKWDAPN 456
 DB 421 KECVNVTVDVPQFNPVQNLTGSAVGQKVTWKWDAPN 456
 RESULT 4
 AAW69490
 ID AAW69490 standard; protein; 456 AA.
 XX
 AC AAW69490;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hgaA, Harep2.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX
 OS Porphyromonas gingivalis.
 XX

PN US5824791-A.
 XX 20-OCT-1998.
 XX 11-DEC-1995; 95US-00570311.
 PR 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX (UYFL) UNIV FLORIDA.
 XX (UABR-) UAB RES FOUND.
 PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
 XX WPI; 1998-582627/49.
 DR N-PSDB; AAV58877.
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease poly.peptide(s)).
 XX (UYFL) UNIV FLORIDA.
 XX (UABR-) UAB RES FOUND.
 PS Claim 1; Col 127-132; 101pp; English.
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX Sequence 456 AA;
 SQ
 Query Match 100.0%; Score 2480; DB 2; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.5e-191;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTPNPNPNPGTITLSEFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICA 60
 DB 1 GTPNPNPNPGTITLSEFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICA 60
 QY 61 SSASYINPEGPONPDNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 DB 61 SSASYINPEGPONPDNLYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 QY 121 FANALLEEVLTAKTVVTAPPAIRGRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
 DB 121 FANALLEEVLTAKTVVTAPPAIRGRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
 QY 181 DDVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGGTN 240
 DB 181 DDVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGGTN 240
 QY 241 VVASFWSNGMALNPNDNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKGTNAGDFTV 300
 DB 241 VVASFWSNGMALNPNDNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKGTNAGDFTV 300
 QY 301 VFEETPNKNGGARFGLSTEADGAKPQSVWIERVTDLPACTKYVAFRHYNCSDLNYILL 360
 DB 301 VFEETPNKNGGARFGLSTEADGAKPQSVWIERVTDLPACTKYVAFRHYNCSDLNYILL 360
 QY 361 DDIOFTMGSSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 DB 361 DDIOFTMGSSPTPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 QY 421 KECVNVTVDVPQFNPVQNLTGSAVGQKVTWKWDAPN 456
 DB 421 KECVNVTVDVPQFNPVQNLTGSAVGQKVTWKWDAPN 456
 RESULT 5
 AAR96030
 ID AAR96030 standard; protein; 2628 AA.
 XX

```

AC AAR96030;
XX
XX 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
XX P. gingivalis haemagglutinin hgaA.
DE
XX
XX Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.
KW
XX
XX Porphyromonas gingivalis; strain 381.
OS
XX
XX Key Location/Qualifiers
FH Peptide 5..21
FT /label= Sig_peptide
XX
XX WO9617936-A2.
PN
XX
XX 13-JUN-1996.
PD
XX
XX 11-DEC-1995; 95WO-US016108.
PF
XX
XX 09-DEC-1994; 94US-00353485.
PR
XX
XX (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
PI
XX
XX WPI; 1996-287181/29.
DR
XX
XX N-PSDB; AAT30654.
DR
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
FT
XX
XX Claim 6; Page 93-101; 153pp; English.
PS
XX
XX P. gingivalis 381 haemagglutinin hgaA (AAR96030) was identified as the
CC product of the hgaA gene (AAT30654) isolated as an EcoRV fragment of
CC genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
CC 24), can be obtd. from transformed host cells and used as a vaccine to
CC protect humans or animals against periodontal disease. Expression in
CC Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
CC also be used to detect the presence of anti-P. gingivalis antibodies and
CC to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
CC 2003 to standardise OS field)
XX
XX
XX Sequence 2628 AA;
SQ
Query Match 100.0%; Score 2480; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 9e-190;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTPNPNPNPGTTLTSESFENGIPASWKITDADGDNNTTTPPGGTSPAGHNSAICA 60
Db 950 GTPNPNPNPGTTLTSESFENGIPASWKITDADGDNNTTTPPGGTSPAGHNSAICA 1009
QY 61 SSASVINFEQPNPNYLVTPELSLPGNGTLTFWVCAQDANYASEHYAVASSTGNDASN 120
Db 1010 SSASVINFEQPNPNYLVTPELSLPGNGTLTFWVCAQDANYASEHYAVASSTGNDASN 1069
QY 121 FANALLEVLTAKTVVTVAPEAIRGRVQGTWYQKTVLPAGTKYVAFRHFGCTDFFWNL 180
Db 1070 FANALLEVLTAKTVVTVAPEAIRGRVQGTWYQKTVLPAGTKYVAFRHFGCTDFFWNL 1129
QY 181 DDVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGCGQWLCSSGQWLTARHGTN 240
Db 1130 DDVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGCGQWLCSSGQWLTARHGTN 1189
QY 241 VVASFSWNGMALPNPNYLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
Db 1190 VVASFSWNGMALPNPNYLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGTNAGDFTV 1249
QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

```

```

Db 1250 VFEETPNGINKGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHYNCSDLNYILL 1309
QY 361 DDIOFTMGGSPTPTDYTYTVYRDGKIKEGITETTFEEDGVATGNHEYCVKVTAGVSP 420
Db 1310 DDIOFTMGGSPTPTDYTYTVYRDGKIKEGITETTFEEDGVATGNHEYCVKVTAGVSP 1369
QY 421 KECVNVTVDPEQFNPNQNLTGSAVGQKVTLLKWDAPN 456
Db 1370 KECVNVTVDPEQFNPNQNLTGSAVGQKVTLLKWDAPN 1405

```

RESULT 6
AAR69488
ID AAR69488 standard; protein; 2628 AA.
XX
AC AAR69488;
XX
DT 22-DEC-1998 (first entry)
XX
DE Haemagglutinin protein hgaA.
XX
KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
XX
OS Porphyromonas gingivalis.
XX
PN US5824791-A.
XX
PD 20-OCT-1998.
XX
PF 11-DEC-1995; 95US-00570311.
PR 08-SEP-1988; 88US-00241640.
PR 25-JAN-1991; 91US-00647119.
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX
XX WPI; 1998-582627/49.
DR N-PSDB; AAV58875.
XX
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s).
XX
PS Claim 1; Col 91-110; 101pp; English.
XX
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hgaA haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 2628 AA;
Query Match 100.0%; Score 2480; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 9e-190;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTPNPNPNPGTTLTSESFENGIPASWKITDADGDNNTTTPPGGTSPAGHNSAICA 60
Db 950 GTPNPNPNPGTTLTSESFENGIPASWKITDADGDNNTTTPPGGTSPAGHNSAICA 1009
QY 61 SSASVINFEQPNPNYLVTPELSLPGNGTLTFWVCAQDANYASEHYAVASSTGNDASN 120
Db 1010 SSASVINFEQPNPNYLVTPELSLPGNGTLTFWVCAQDANYASEHYAVASSTGNDASN 1069
QY 121 FANALLEVLTAKTVVTVAPEAIRGRVQGTWYQKTVLPAGTKYVAFRHFGCTDFFWNL 180
Db 1070 FANALLEVLTAKTVVTVAPEAIRGRVQGTWYQKTVLPAGTKYVAFRHFGCTDFFWNL 1129
QY 181 DDVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGCGQWLCSSGQWLTARHGTN 240
Db 1130 DDVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGCGQWLCSSGQWLTARHGTN 1189
QY 241 VVASFSWNGMALPNPNYLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
Db 1190 VVASFSWNGMALPNPNYLISKDVTGATKVKYIYAVNDGPPGDHYAVMISKTGTNAGDFTV 1249
QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHYNCSDLNYILL 1369

QY 181 DDVEIKANGKRADETFETFESSHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGCTN 240
 DB 1130 DDVEIKANGKRADETFETFESSHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGCTN 1189
 QY 241 VVASFSWNGMALNPNDNLIISKDVTGATKVKYKYAVNDGFFPDGHYAVMISKTGTNAGDFTV 300
 DB 1190 VVASFSWNGMALNPNDNLIISKDVTGATKVKYKYAVNDGFFPDGHYAVMISKTGTNAGDFTV 1249
 QY 301 VFETPTNGINKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHHYNCSDLNKYL 360
 DB 1250 VFETPTNGINKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHHYNCSDLNKYL 1309
 QY 361 DDIOFTMGSSPTDYTYVYRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 DB 1310 DDIOFTMGSSPTDYTYVYRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 1369
 QY 421 KECVNVTVDPVQFNPNQNLTSAGVGQKVTWKWDAPN 456
 DB 1370 KECVNVTVDPVQFNPNQNLTSAGVGQKVTWKWDAPN 1405

RESULT 7

AAR96021
 ID AAR96021 standard; protein; 450 AA.

AC AAR96021;

XX 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX P. gingivalis haemagglutinin hAgA Harepl product.

DE Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harepl.

XX Porphyromonas gingivalis; strain 381.

XX MO9617936-A2.

XX 13-JUN-1996.

XX 11-DEC-1995; 95WO-US016108.

XX 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX WPI; 1996-287181/29.

DR N-PSDB; AAT30645.

XX Porphyromonas gingivalis genes and proteins - used in the detection and
 vaccination against periodontal disease.

XX Claim 4; Page 103-104; 153pp; English.

XX Harepl (AAR96021) is the product of the Harepl repeat unit (AAT30645) of
 the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
 haemagglutinin hAgA (see also AAR96030). Harepl and other hAgA repeat
 unit products (see also AAR96022-24) can be obtd. from transformed host
 cells and used as vaccines to protect humans or animals against
 periodontal disease. Expression in Salmonella cells allows prodn. of live
 vaccine. Harepl-4 can also be used to detect the presence of anti-p.
 gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 appln. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 450 AA;

Query Match 98.2%; Score 2436; DB 2; Length 450;

Best Local Similarity 99.8%; Pref. No. 2.6e-187;

Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 PNPNGTTTLLSEFENGIPASWKTIDADGGNNWTTTTPPGGTSFAGHNSAICASSASYI 66
 DB 1 PNPNGTTTLLSEFENGIPASWKTIDADGGNNWTTTTPPGGTSFAGHNSAICASSASYI 60
 QY 67 NFEQPNDNLYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
 DB 61 NFEQPNDNLYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
 QY 127 EVULTAKTVVTAPEAIRGTRVQGTWYOKTVOLPAGTKYVAFRHHGCTDFFWNLDDVEIK 186
 DB 121 EVULTAKTVVTAPEAIRGTRVQGTWYOKTVOLPAGTKYVAFRHHGCTDFFWNLDDVEIK 180
 QY 187 ANGKRADFTTFESSSTHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGCTNVAFS 246
 DB 181 ANGKRADFTTFESSSTHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGCTNVAFS 240
 QY 247 WNGMALNPNDNLIISKVGTGATKVKYKYAVNDGFFPDGHYAVMISKTGTNAGDFTVFEETP 306
 DB 241 WNGMALNPNDNLIISKVGTGATKVKYKYAVNDGFFPDGHYAVMISKTGTNAGDFTVFEETP 300
 QY 307 NGINKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHHYNCSDLNKYLDDIOFT 366
 DB 301 NGINKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHHYNCSDLNKYLDDIOFT 360
 QY 367 MGSPTPTDYTYVYRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVN 426
 DB 361 MGSPTPTDYTYVYRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVN 420
 QY 427 TVDPVQFNPNQNLTSAGVGQKVTWKWDAPN 456
 DB 421 TVDPVQFNPNQNLTSAGVGQKVTWKWDAPN 450

RESULT 8

AAR69489

ID AAR69489 standard; protein; 450 AA.

AC AAR69489;

XX 22-DEC-1998 (first entry)

DT Haemagglutinin protein hAgA, Harepl.

DE Haemagglutinin protein; periodontal disease; vaccine; hAgA.

XX Porphyromonas gingivalis.

OS US5824791-A.

XX 20-OCT-1998.

XX 11-DEC-1995; 95US-00570311.

XX 08-SEP-1988; 88US-00241640.

XX 25-JAN-1991; 91US-00647119.

XX 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;

XX WPI; 1998-582627/49.

DR N-PSDB; AAV58876.

XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 protease poly(peptide(s)).

XX Claim 1; Col 121-126; 101pp; English.

XX This sequence is encoded by a Porphyromonas gingivalis gene of the

CC invention. This sequence represents the hAgA haemagglutinin protein. The

CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease

XX
 SQ Sequence 450 AA;

Query Match 98.2%; Score 2436; DB 2; Length 450;
 Best Local Similarity 99.8%; Pred. No. 2.6e-187;
 Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 PNPNGTTLSEFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASSASYI 66
 Db 1 PNPNGTTLSEFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASSASYI 60
 QY 67 NFGPQNPQNYLVTPELSLPGGTLTFWCAQDANYASVHYVYASSTGNDASNFANALL 126
 Db 61 NFGPQNPQNYLVTPELSLPGGTLTFWCAQDANYASVHYVYASSTGNDASNFANALL 120
 QY 127 EEVLTAKTVVTAPALRGTRVQGTWKIVQLPAGTKYVAFRHFCTDFFWNLDDVEIK 186
 Db 121 EEVLTAKTVVTAPALRGTRVQGTWKIVQLPAGTKYVAFRHFCTDFFWNLDDVEIK 180
 QY 187 ANGKRADEFETPSSSTHGEAPAEWTTIDADGQGWCLSSGQDGLWLTAGGTTNVVASFS 246
 Db 181 ANGKRADEFETPSSSTHGEAPAEWTTIDADGQGWCLSSGQDGLWLTAGGTTNVVASFS 240
 QY 247 WNGMALNPQNYLISKDVATKVKYVYVNDGPPGDHYAVMISKTGTNAGDFTVVEETP 306
 Db 241 WNGMALNPQNYLISKDVATKVKYVYVNDGPPGDHYAVMISKTGTNAGDFTVVEETP 300
 QY 307 NGINKGARFGLSTEADGAKPQSVWIERVTDLPAKTGYVAFRHYNGSDLNILLDDIQT 366
 Db 301 NGINKGARFGLSTEADGAKPQSVWIERVTDLPAKTGYVAFRHYNGSDLNILLDDIQT 360
 QY 367 MGSSTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVNV 426
 Db 361 MGSSTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKCVNV 420
 QY 427 TVDPVQFNQVQNLTGSAVGQKVTWKWDAFN 456
 Db 421 TVDPVQFNQVQNLTGSAVGQKVTWKWDAFN 450

RESULT 9
 AAR96024
 ID AAR96024 standard; protein; 439 AA.

XX AAR96024;
 AC AAR96024;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hAgA Harep4 product.
 XX
 KW Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harep4.

XX Porphyromonas gingivalis; strain 381.

OS AAR96024 standard; protein; 439 AA.

XX WO9617936-A2.

XX 13-JUN-1996.

PF 11-DEC-1995; 95WO-US016108.

XX 09-DEC-1994; 94US-00353485.

XX (UYEL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

PI Progluske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX WPI; 1996-287181/29.

DR N-PSDB; AAT30648.

XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.

XX Claim 4; Page 114-115; 153pp; English.

XX Harep4 (AAR96024) is the product of the Harep4 repeat unit (AAT30648) of
 CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
 CC haemagglutinin hAgA (see also AAR96030). Harep4 and other hAgA repeat
 CC unit products (see also AAR96021-23) can be obt'd. from transformed host
 CC cells and used as vaccines to protect humans or animals against
 CC periodontal disease. Expression in Salmonella cells allows prodn. of live
 CC vaccine Harep4-4 can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)

XX SQ Sequence 439 AA;

Query Match 94.8%; Score 2351; DB 2; Length 439;
 Best Local Similarity 98.2%; Pred. No. 1.7e-180;
 Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTFNPNPNPQNTTTLSEFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 60
 Db 1 GTFNPNPNPQNTTTLSEFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICV 60
 QY 61 SSASINFEPPQNPQNYLVTPELSLPGGTLTFWCAQDANYASVHYVYASSTGNDASN 120
 Db 61 SSASINFEPPQNPQNYLVTPELSLPGGTLTFWCAQDANYASVHYVYASSTGNDASN 120
 QY 121 FANALLEEVLTAKTVVTAPALRGTRVQGTWKIVQLPAGTKYVAFRHFCTDFFWNL 180
 Db 121 FANALLEEVLTAKTVVTAPALRGTRVQGTWKIVQLPAGTKYVAFRHFCTDFFWNL 180
 QY 181 DDVEIKANGKRADEFETPSSSTHGEAPAEWTTIDADGQGWCLSSGQDGLWLTAGGTTN 240
 Db 181 DDVEIKANGKRADEFETPSSSTHGEAPAEWTTIDADGQGWCLSSGQDGLWLTAGGTTN 240
 QY 241 VVASFSWNGMALNPQNYLISKDVATKVKYVYVNDGPPGDHYAVMISKTGTNAGDFTV 300
 Db 241 VVASFSWNGMALNPQNYLISKDVATKVKYVYVNDGPPGDHYAVMISKTGTNAGDFTV 300
 QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERVTDLPAKTGYVAFRHYNGSDLNILL 360
 Db 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERVTDLPAKTGYVAFRHYNGSDLNILL 360
 QY 361 DDIOFTMGGSPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 Db 361 DDIOFTMGGSPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 QY 421 KECVNVTVDPVQFNQVQNL 439
 Db 421 KCVNVNTINPTQFNQVQNL 439

RESULT 10

AAR69492
 ID AAR69492 standard; protein; 439 AA.

XX AAR69492;

XX 22-DEC-1998 (first entry)

DE Haemagglutinin protein hAgA, Harep4.

XX Haemagglutinin protein; periodontal disease; vaccine; hAgA.

OS Porphyromonas gingivalis.

XX US5824791-A.

XX 20-OCT-1998.

XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulskie-Fox A, Lepine G;
XX WPI; 1998-582627/49.
XX N-PSDB; AAV58879.
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly(peptide(s)).
XX Claim 1; Col 139-144; 101pp; English.
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX invention. This sequence represents the hga haemagglutinin protein. The
XX polypeptides are used to produce antibodies to organisms associated with
XX periodontal disease. The antibodies are also used in purification and
XX identification procedures. The genes and polypeptides are used as
XX vaccines against periodontal disease
XX
XX Sequence 439 AA;
Query Match 94.8%; Score 2351; DB 2; Length 439;
Best Local Similarity 98.2%; Pred. No. 1.7e-180;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 GTPNPNPNPGTTLTSEFENGIPASWKTIDADGDNWTTTPPGGTSPAGHNSAICA 60
Db 1 GTPNPNPNPGTTLTSEFENGIPASWKTIDADGDNWTTTPPGGTSPAGHNSAICV 60
QY 61 SSASVINEFGQPNPNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 61 SSASVINEFGQPNPNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEVLTAKTVVTAPAIRGRVGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 180
Db 121 FANALLEVLTAKTVVTAPAIRGRVGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 180
QY 181 DDVEIKANGKADFTETPESSTHGEAPAEWTTIDADGGQGWLCSSQQLGWLTAHGGTN 240
Db 181 DDVEIKANGKADFTETPESSTHGEAPAEWTTIDADGGQGWLCSSQQLGWLTAHGGTN 240
QY 241 VVAFSWMNGMALPNPNYLISKDVTGATKVKYKYAVNDGFGPDGHYAVMI SKGTGNAGDTV 300
Db 241 VVAFSWMNGMALPNPNYLISKDVTGATKVKYKYAVNDGFGPDGHYAVMI SKGTGNAGDTV 300
QY 301 VFEETPNGINKGARFGLSTADGAKPOSWMIERTVDLPAGTKYVAFRHYNCSDILNYLL 360
Db 301 VFEETPNGINKGARFGLSTADGAKPOSWMIERTVDLPAGTKYVAFRHYNCSDILNYLL 360
QY 361 DDIOFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGATGNHHCYEVKVTAGVSP 420
Db 361 DDIOFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGATGNHHCYEVKVTAGVSP 420
QY 421 KECVNTVDPVQFNPQNL 439
Db 421 KVCVNTVNPQFNPQNL 439
RESULT 11
ID AAR96033 standard; protein; 1687 AA.
XX AAR96033;
AC AAR96033;
XX 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)
XX P. gingivalis haemagglutinin hage.
XX Haemagglutinin; hage; periodontal disease; vaccine; antibody.
XX Porphyromonas gingivalis; strain FDC381.
XX WO9617936-A2.
XX 13-JUN-1996.
XX 11-DEC-1995; 95WO-US016108.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX Progulskie-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX WPI; 1996-287181/29.
XX N-PSDB; AAT30656.
XX Porphyromonas gingivalis genes and proteins - used in the detection and
XX vaccination against periodontal disease.
XX Claim 5; Page 138-143; 153pp; English.
XX P. gingivalis 381 haemagglutinin hage (AAR96033) was identified as the
XX product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
XX The haemagglutinin can be obt'd. from transformed host cells and used as a
XX vaccine to protect humans or animals against periodontal disease.
XX Expression in Salmonella cells allows prodn. of a live vaccine. The
XX haemagglutinin can also be used to detect the presence of anti-P.
XX gingivalis antibodies and to raise monoclonal antibodies for diagnostic
XX appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 1687 AA;
Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 5.2e-159;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;
QY 3 PNPENPNPGTTLTSEFENGIPASWKTIDADGDNWTTTPPGGTSPAGHNSAICA 62
Db 935 PNPENPNPGTTLTSEFENGIPASWKTIDADGDNWTTTPPGGTSPAGHNSAICV 991
QY 63 ASY-INFEQPNPNYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 121
Db 992 ESFGLGGIGVLTDPNYLTTPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASN 1051
QY 122 ANALLEVLTAKTVVTAPAIRGRVGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 181
Db 1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFCTDFFWNL 1109
QY 182 DVEIKANGKADFTETPESSTHGEAPAEWTTIDADGGQGWLCSSQQLGWLTAHGGTN 241
Db 1110 EVEIKANGKADFTETPESSTHGEAPAEWTTIDADGGQGWLCSSQQLGWLTAHGGTN 1169
QY 242 VVAFSWMNGMALPNPNYLISKDVTGATKVKYKYAVNDGFGPDGHYAVMI SKGTGNAGDTV 301
Db 1170 VVAFSWMNGMALPNPNYLISKDVTGATKVKYKYAVNDGFGPDGHYAVMI SKGTGNAGDTV 1229
QY 302 FEETPNGINKGARFGLSTADGAKPOSWMIERTVDLPAGTKYVAFRHYNCSDILNYLL 361
Db 1230 FEETPNGINKGARFGLSTADGAKPOSWMIERTVDLPAGTKYVAFRHYNCSDILNYLL 1289
QY 362 DDIOFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGATGNHHCYEVKVTAGVSP 421
Db 1290 DDIOFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGATGNHHCYEVKVTAGVSP 1349
QY 422 ECVNTVDPVQFNPQNLTGSVAGQKVTLLKWDAPN 456

Query Match 84.0%; Score 2082; DB 2; Length 1704;
 Best Local Similarity 86.2%; Pred. No. 5.2e-158;
 Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 63 ASY- INFEGPQNPNDNYLTPELSLPNGGTLTFWCAQDANYASHEHYAVASSTGNDASNF 121
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 1009 ESFGLGGIGVLTDPNYLITPALDLPGGKLTFFWCAQDANYASHEHYAVASSTGNDASNF 1068
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 122 ANALLEEVLTAKTVVTAPEAIRGTVOGTWYQKTVQVLPAGTKYVAFRHFQCTDFFWINLD 181
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAFTKYVAFRHFQSTDMFYDLD 1126
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 182 DVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 241
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 1127 EVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 1186
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 242 VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 301
 DB VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 301
 QY 1187 VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 1246
 DB VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 1246
 QY 302 FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 361
 DB FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 361
 QY 1247 FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 1306
 DB FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 1306
 QY 362 DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 421
 DB DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 421
 QY 1307 DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 1366
 DB DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 1366
 QY 422 ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 456
 DB ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 456
 QY 1367 ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 1401
 DB ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 1401

RESULT 14
 AA034843
 ID AA034843 standard; protein; 1704 AA.
 AC AA034843;
 XX AA034843;
 DT 03-JUN-1998 (first entry)
 XX Arg-gingipain high molecular weight prepolyprotein sequence.
 DE Arg-specific gingipain protease; gingivalis; periodontal disease;
 KW vaccine; infection.
 XX Porphyromonas gingivalis.
 OS Porphyromonas gingivalis.

Key Location/Qualifiers
 FT Protein 1..227
 FT /note= "precursor protein"
 XX
 XX WO9734629-A1.
 XX
 XX 25-SEP-1997.
 XX
 XX 21-MAR-1997; 97WO-US004635.
 XX
 XX 22-MAR-1996; 96US-0013945P.
 XX
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA (MORE-) MOREHOUSE SCHOOL MEDICINE.
 XX
 XX Potempa J, Travis J, Genco C;
 PI WPI; 1997-479993/44.
 XX N-PSDB; AAT93872.
 DR
 XX

PT Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
 PT useful for protecting animals and humans from gingivalis and periodontal
 PT diseases.
 XX Disclosure; Page 68-73; 95pp; English.
 PS The present sequence represents an arginine-specific protease of
 CC Porphyromonas gingivalis. The following peptides, derived from Arg- and
 CC Lys-specific high molecular weight proteases, offer protection against
 CC infection: YTVTRDQK IREGTATTE DDGVTAGNHE YCEKVTAGS VSPKVC (I);
 CC YTPVEEKQNG RMIVIVAKKY (II); QLPFIQDVAC VNGDFLEFSP CFBAALMRAQ (III);
 CC GEPNYPQVS NITATTQGGK VTLKWDAPSTK (IV); GNEHYCEVVK YTAGVSPKVC KDVTY (V);
 CC RMFNPYEPGR YTPVEEKQNG (VI); TFAGFEDTYK RMFNPYEPGR (VII); DYTIVYRDG
 CC TKIKEGLTAT TEEDGVATG NMEYCVKVCYK TAGSPKVC (VIII); YTVYRDGT KIKEGLTATTF
 CC EEDG (IX); RGTGKKEGL TATTEEDGV ATGN (X); KIKEGLTATTF FEDGVTAGN HEY (XI)
 CC ; KWDAPNGTPN PNPEN PNPEN PETITLSE (XII); and YTPVEEKNG RMIVIVAKKY
 CC (XIII). They are used in vaccines to protect animals, including humans,
 CC from gingivitis and/or periodontal diseases
 XX Sequence 1704 AA;
 SQ

Query Match 84.0%; Score 2082; DB 2; Length 1704;
 Best Local Similarity 86.2%; Pred. No. 5.2e-158;
 Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 63 ASY- INFEGPQNPNDNYLTPELSLPNGGTLTFWCAQDANYASHEHYAVASSTGNDASNF 121
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 1009 ESFGLGGIGVLTDPNYLITPALDLPGGKLTFFWCAQDANYASHEHYAVASSTGNDASNF 1068
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 122 ANALLEEVLTAKTVVTAPEAIRGTVOGTWYQKTVQVLPAGTKYVAFRHFQCTDFFWINLD 181
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAFTKYVAFRHFQSTDMFYDLD 1126
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 182 DVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 241
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 1127 EVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 1186
 DB ENPNPNPNPGTTLTSESFENGIPASWKIIDADGDNWTTTPPGGTSPAGHNSAICASS 62
 QY 242 VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 301
 DB VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 301
 QY 1187 VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 1246
 DB VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKTGNAGDFTVV 1246
 QY 302 FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 361
 DB FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 361
 QY 1247 FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 1306
 DB FEETPENGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHFQCTDFFWINLD 1306
 QY 362 DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 421
 DB DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 421
 QY 1307 DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 1366
 DB DIQFTMGSSPTDITYTVYRDGDKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 1366
 QY 422 ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 456
 DB ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 456
 QY 1367 ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 1401
 DB ECNVTVDPVQFNPVQNLGTSAGVQKVTWKWDAPN 1401

RESULT 15
 AA067396
 ID AA067396 standard; protein; 1704 AA.
 AC AA067396;
 XX AA067396;
 DT 25-APR-2000 (first entry)
 XX Arg-gingipain-2 amino acid sequence.
 DE Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
 KW immunogenic component; vaccine; inflammatory response; tissue damage;
 KW periodontal disease.
 XX Porphyromonas gingivalis.
 OS

```
XX Key Location/Qualifiers
FH Region 229..719
FT /note="Amino acids 229-719 are specifically claimed"
FT Region 720..1185
FT /note="Amino acids 720-1185 are specifically claimed"
XX
PN US6017532-A.
XX
XX 25-JAN-2000.
XX
XX 08-NOV-1994; 94US-00336308.
XX
XX 10-SEP-1993; 93US-00119361.
XX 24-JUN-1994; 94US-00265441.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX Potempa JS, Travis J;
XX
XX WPI: 2000-136659/12.
XX N-PSDB; AA260181.
XX
XX New Porphyromonas gingivalis arginine-specific protease preparation
PT useful for preparing vaccines against periodontal disease and for
PT screening for Arg-gingipain inhibitors.
XX
XX Claim 1; Col 29-42; 55pp; English.
XX
XX This sequence represents a Porphyromonas gingivalis arginine-specific
CC proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC Gingipain-2 consists of a 50kD protease component non-covalently
CC associated with a 44kD haemagglutinin component. The proteinase is
CC stimulated by glycine containing peptides and glycine analogues. It is
CC inhibited by cysteine protease group specific inhibitors. The protease
CC preparation can be used in immunogenic compositions and vaccines against
CC inflammatory response and tissue damage caused by P. gingivalis in
CC periodontal disease. It can also be used to screen for agents that
CC modulate Arg-gingipain proteinase activity inhibitors
XX
XX Sequence 1704 AA;
XX
Query Match 84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred No. 5.2e-158;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;
QY 3 PNPENPNPPTTLSESPENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAICASS 62
Db 952 PNPENPNPPTTLSESPENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAICASS 62
QY 63 ASY-INPFGPQPNPDNLYVTPELSLPGGTLTFWVCAQDANYASEHYAYVASTGNDASNF 121
Db 1009 ESFGLGGIGVLTPDNLITPDLDPNGGKLTFWVCAQDANYASEHYAYVASTGNDASNF 1068
QY 122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWQKTVQLPAGTKYVAFRHFQCTDPPWNLID 181
Db 1069 TNALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVQLPAGTKYVAFRHFQCTDPPWNLID 1126
QY 182 DVEIKANGKRADFTTFSSSTHGEAPAEWTTIDADGGQGMCLSSGQLGMLTAHGGTNV 241
Db 1127 EYEIKANGKRADFTTFSSSTHGEAPAEWTTIDADGGQGMCLSSGQLGMLTAHGGTNV 1186
QY 242 VASFSWNGMALNPNDNLYLSKQVTGATKYKYVAVNDGPPGDHYAVMISKTGTNAGDFTVV 301
Db 1187 VASFSWNGMALNPNDNLYLSKQVTGATKYKYVAVNDGPPGDHYAVMISKTGTNAGDFTVV 1246
QY 302 FEETPNGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNLYLLD 361
Db 1247 FEETPNGKKGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNLYLLD 1306
QY 362 DIQFTMGSSPTPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSPK 421
Db 1307 DIQFTMGSSPTPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSPK 1366
```

QY 422 ECVNVTVDVPQFNPVQNLITGSAVGOKVTLKWDAPN 456
Db 1367 ECVNVTINFTQNPVKNLKAQPDGDDVVLKWEAPS 1401

Search completed: May 18, 2004, 11:42:42
Job time : 49.0287 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:37:00 ; Search time 11.906 Seconds
(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTPNPNPNPGTTTSLSESF.....QNLGSAVGQKVKTLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 2480 | 100.0 | 2628 | 2 | hemagglutinin A - |
| 2 | 2082 | 84.0 | 1704 | 2 | gingipain R (EC 3. |
| 3 | 2059 | 83.0 | 1526 | 2 | gingipain R (EC 3. |
| 4 | 2034.5 | 82.0 | 1732 | 2 | lysine-specific cy |
| 5 | 266 | 10.7 | 991 | 2 | arginyl endopeptid |
| 6 | 139.5 | 5.6 | 1052 | 2 | conserved hypothet |
| 7 | 139.5 | 5.6 | 1341 | 2 | hypothetical prote |
| 8 | 132.5 | 5.3 | 691 | 2 | hypothetical prote |
| 9 | 131.5 | 5.3 | 2468 | 2 | hypothetical prote |
| 10 | 128.5 | 5.2 | 1248 | 2 | autolysin (impor |
| 11 | 128 | 5.2 | 4936 | 2 | hypothetical prote |
| 12 | 127.5 | 5.1 | 1684 | 2 | amylase A-180 - al |
| 13 | 127 | 5.1 | 713 | 2 | hypothetical prote |
| 14 | 126 | 5.1 | 1441 | 2 | phage pil prote |
| 15 | 126 | 5.1 | 1649 | 2 | hypothetical prote |
| 16 | 124.5 | 5.0 | 1904 | 2 | tail-host specific |
| 17 | 124.5 | 5.0 | 5291 | 2 | hypothetical prote |
| 18 | 123.5 | 5.0 | 5188 | 2 | probable RTX fami |
| 19 | 122.5 | 4.9 | 1274 | 2 | transferrin-like p |
| 20 | 121.5 | 4.9 | 465 | 2 | S-layer protein - |
| 21 | 121.5 | 4.9 | 3624 | 2 | large repetitive p |
| 22 | 121 | 4.9 | 1090 | 2 | cellulose 1,4-beta |
| 23 | 120 | 4.8 | 908 | 2 | hypothetical prote |
| 24 | 120 | 4.8 | 2783 | 2 | hypothetical prote |
| 25 | 119 | 4.8 | 715 | 2 | alkaline serine pr |
| 26 | 119 | 4.8 | 1873 | 2 | surface protein pr |
| 27 | 118.5 | 4.8 | 926 | 2 | hypothetical prote |
| 28 | 118.5 | 4.8 | 1385 | 2 | parasporal crystal |
| 29 | 118.5 | 4.8 | 2817 | 2 | uncharacterized pr |

30 118.5 4.8 4199 2 S76412
31 118 4.8 1461 2 E90696
32 118 4.8 1461 2 A85547
33 117.5 4.7 1034 2 T30551
34 116.5 4.7 1635 2 A10452
35 115.5 4.7 702 2 S48753
36 115.5 4.7 875 2 AF0472
37 115.5 4.7 1118 1 A49724
38 115.5 4.7 1282 2 UC4393
39 115.5 4.7 1939 2 D97316
40 115 4.6 821 2 AD1507
41 115 4.6 1651 2 JC1340
42 115 4.6 3029 2 S76109
43 114 4.6 872 2 S49541
44 113 4.6 1345 2 H90975
45 112.5 4.5 938 2 AFI772

ALIGNMENTS

RESULT 1

T28651 hemagglutinin A - Porphyromonas gingivalis
C:Species: Porphyromonas gingivalis
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C;Accession: T28651
R;Han, N.; Whitlock, J.; Progulskie-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A;Title: The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381 contains four 1
A;Reference number: Z20494; MUID:97047672; PMID:8926061
A;Accession: T28651
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2628 <HAN>
A;Cross-references: EMBL:U41807; NID:g1552410; PID:g1469916; PIDN:AA817128.1
C;Genetics:
A;Gene: haga

Query Match 100.0%; Score 2480; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.4e-162;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTPNPNPNPGTTTSLSESPENGIPASWKTIADGNGNNWTTTPPGTGSFAGHNSAICA 60
Db 950 GTPNPNPNPGTTTSLSESPENGIPASWKTIADGNGNNWTTTPPGTGSFAGHNSAICA 1009
Qy 61 SSASYINFEQPNDPNVLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 1010 SSASYINFEQPNDPNVLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069
Qy 121 FANALLEEVLTAKTAVTAPAIRGTVOGTWYQKTVQLPAGTKYVAFRHHGCTDFFWINL 180
Db 1070 FANALLEEVLTAKTAVTAPAIRGTVOGTWYQKTVQLPAGTKYVAFRHHGCTDFFWINL 1129
Qy 181 DDVEIKANGKRAUFTETFESSSTHGEAPAEWTTTIDADGGGWLCLSSGQLGWLTAHGCTN 240
Db 1130 DDVEIKANGKRAUFTETFESSSTHGEAPAEWTTTIDADGGGWLCLSSGQLGWLTAHGCTN 1189
Qy 241 VVASFNGMALNPDNLYISKDVTKYKYAVNDGFGPDHYAVMISKTGTNAGDFTV 300
Db 1190 VVASFNGMALNPDNLYISKDVTKYKYAVNDGFGPDHYAVMISKTGTNAGDFTV 1249
Qy 301 VFEETPNGKKGARGLSTEADGAKPQSVWITRTVDLPAGTKYVAFRHHNSDLNLYILL 360
Db 1250 VFEETPNGKKGARGLSTEADGAKPQSVWITRTVDLPAGTKYVAFRHHNSDLNLYILL 1309
Qy 361 DDIOFTWGSPTPTDYTYTVYRDGTIKGLTETTFEEDGATGNHEYCVVEKYTAGVSP 420
Db 1310 DDIOFTWGSPTPTDYTYTVYRDGTIKGLTETTFEEDGATGNHEYCVVEKYTAGVSP 1369
Qy 421 KECNVTVDPQNPVQNLTGSAVGQKVKTLKWDAPN 456
|||||

Db 1370 KCVNVTVDVQPNVQNLTSAGVQKVLKWDAPN 1405

RESULT 2

A55426

gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis

N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; R

C;Species: Porphyromonas gingivalis

C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999

C;Accession: A55426; D53113

R;Payloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, H

J. Biol. Chem. 270, 1007-1010, 1995

A;Title: Molecular cloning and structural characterization of the Arg-gingipain proteinase

A;Reference number: A55426; MUID:95138080; PMID:7836351

A;Accession: A55426

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1704 <PAV>

A;Cross-references: GB:U15282; NID:G557067; PIDN:AAA69539.1; PID:G557068

R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.

J. Biol. Chem. 269, 406-411, 1994

A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat

A;Reference number: A53113; MUID:94103245; PMID:8276827

A;Accession: D53113

A;Status: preliminary

A;Molecule type: protein

A;Residues: 228-249 <PIK>

A;Experimental source: H66

A;Note: sequence extracted from NCBI backbone (NCBIP:141694)

C;Keywords: cysteine proteinase; hydrolase

Query Match 84.0%; Score 2082; DB 2; Length 1704;

Best Local Similarity 86.2%; Pred. No. 2.2e-135;

Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTTIDADGGNNWTTTPPGTSPAGHNSAICASS 62

Db 952 PNPENPNPGTTLTSESFENGIPASWKTTIDADGGHGWKPGNAPG---IAGYNSGCVYS 1008

QY 63 ASY-INFEQPNPNVLPPELSPNGGTLTFWVCAQADANYASEHYAVYASSTGNDASNF 121

Db 1009 ESFGLGGIGVLTDPDNLITPALDLPNGGKLTFWVCAQADANYASEHYAVYASSTGNDASNF 1068

QY 122 ANALLEEVLTAKTIVTAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFQCTDFFWINDL 181

Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLD 1126

QY 182 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGGTIV 241

Db 1127 EVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGGTIV 1186

QY 362 DIQFTMGGSPTPTDYTYTVYRDGTKIKGLTETTFEEDGVATGNHEYCEVEKYTAGVSPK 421

Db 1307 DIQFTMGGSPTPTDYTYTVYRDGTKIKGLTETTFEEDGVATGNHEYCEVEKYTAGVSPK 1366

QY 422 ECNVTVDPVQPNVQNLTSAGVQKVLKWDAPN 456

Db 1367 KCVNVTINPTQPNVKNLKAQPDGGDVVLKWEAPS 1401

RESULT 3

S49763

gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)

C;Species: Porphyromonas gingivalis

C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C;Accession: S49763

R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajah, M.; Curtis, M.A.

submitted to the EMBL Data Library, November 1994

A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 of

A;Reference number: S49763

C;Accession: S49763

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1526 <ADU>

A;Cross-references: EMBL:X82680

C;Genetics:

A;Gene: prpR1

C;Keywords: cysteine proteinase; hydrolase

Query Match 83.0%; Score 2059; DB 2; Length 1526;

Best Local Similarity 85.5%; Pred. No. 7.4e-134;

Matches 389; Conservative 17; Mismatches 43; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTTIDADGGNNWTTTPPGTSPAGHNSAICASS 62

Db 954 PNPENPNPGTTLTSESFENGIPASWKTTIDADGGHGWKPGNAPG---IAGYNSGCVYS 1010

QY 63 ASY-INFEQPNPNVLPPELSPNGGTLTFWVCAQADANYASEHYAVYASSTGNDASNF 121

Db 1011 ESFGLGGIGVLTDPDNLITPALDLPNGGKLTFWVCAQADANYASEHYAVYASSTGNDASNF 1070

QY 122 ANALLEEVLTAKTIVTAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFQCTDFFWINDL 181

Db 1071 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLD 1128

QY 182 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGGTIV 241

Db 1129 EVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGGTIV 1188

QY 242 VASFSWNGMALPNPDNLYISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 301

Db 1189 VSFSGWNGMALPNPDNLYISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 1248

QY 302 FEETPNGINKGARGFLSTEADGAKPQSVWIERTVDLDPAGTKYVAFRHYNSCDLNYILLD 361

Db 1249 FEETPNGINKGARGFLSTEADGAKPQSVWIERTVDLDPAGTKYVAFRHYNSCDLNYILLD 1308

QY 362 DIQFTMGGSPTPTDYTYTVYRDGTKIKGLTETTFEEDGVATGNHEYCEVEKYTAGVSPK 421

Db 1309 DIQFTMGGSPTPTDYTYTVYRDGTKIKGLTETTFEEDGVATGNHEYCEVEKYTAGVSPK 1368

QY 422 ECNVTVDPVQPNVQNLTSAGVQKVLKWDAPN 456

Db 1369 KCVNVTINPTQPNVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 4

T30836

lysine-specific cysteine proteinase porphyrain (EC 3.4.22.-) - Porphyromonas gingivalis

N;Alternate names: lysine-specific cysteine proteinase 1, 60K

C;Species: Porphyromonas gingivalis

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000

C;Accession: T30836; T30837; T30526; A53113

R;Barkocy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulski-Fox, A.; Lantz

J. Bacteriol. 178, 2734-2741, 1996

A;Title: Analysis of the prtp gene encoding porphyrain, a cysteine proteinase of Porphy

A;Reference number: Z20895; MUID:96213011; PMID:8631659

A;Accession: T30836

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-1732 <BAR>

A;Cross-references: EMBL:U42210; NID:gl314325; PID:gl314326; PIDN:AA06565.1

R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z20896

A;Accession: T30837

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A;Cross-references: EMBL:U075366; NID:G2182811; PID:G2182812; PIDN:AA60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in *Porphyromonas gingivalis*. Isolation and characterization of the IS195.
A;Reference number: 220844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:G2738802; PID:G2738803; PIDN:AAC26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from *Porphyromonas gingivalis*. Isolation and characterization of the proteinases.
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match 82.0%; Score 2034.5; DB 2; Length 1732;
Best Local Similarity 84.3%; Pred. No. 4.3e-132;
Matches 387; Conservative 19; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPPT-TTISESPENGIPASWKITIDAGDGNWTTTPPGGTSFAGHNSAIC 59
Db 969 GTPNPNPNPNPTTISESPENGIPASWKITIDAGDGHGKPNAPG---IAGYNSGC 1025

QY 60 ASSASY-INFEQPNPNLYLTPSLPNSGTLTFWCAQDANYASEHYAVASSTGND 118
Db 1026 VYSESGGLGGIGVLTDPNLYLTPALDLPNGCKLTFWCAQDANYASEHYAVASSTGND 1085

QY 119 SNFANALLEEVLTAKTVTAPEARTGTRVQGTWQKTVQAGTKYVAFRHFSGTDFWI 178
Db 1086 SNFTNALLEETITAKG-VRSFKAIRG-RIQGTWRQKTVQAGTKYVAFRHFQSTDMFYI 1143

QY 179 NLDDVEIKANGKRAFDTETPSSPHGAPAEWTIDAGDGGQWCLSSGOLGWLTAHGG 238
Db 1144 DLDEVEIKANGKRAFDTETPSSPHGAPAEWTIDAGDGGQWCLSSGOLGWLTAHGG 1203

QY 239 TNVAVSFWNGMALPNLYLSKDVTKATKYKYAVNDGPPGPHYAVMISKTGTNAGDF 298
Db 1204 SNVVSFSFWNGMALPNLYLSKDVTKATKYKYAVNDGPPGPHYAVMISKTGTNAGDF 1263

QY 299 TVVFEETPNKNGKARFGLSTADGAKPQSVWIERTVQAGTKYVAFRHYNGSDLYNI 358
Db 1264 TVVFEETPNKNGKARFGLSTADGAKPQSVWIERTVQAGTKYVAFRHYNGSDLYNI 1323

QY 359 LLDDIQFTMGSPPTDPTTYTVVVDGDKIKEGLTETTFEEDGVATGNHCEVVEKYTAGV 418
Db 1324 LLDDIQFTMGSPPTDPTTYTVVVDGDKIKEGLTETTFEEDGVATGNHCEVVEKYTAGV 1383

QY 419 SPKECVNVTVDPQVNPQNLTGSV--GQKVTLKWDAP 455
Db 1384 SPKKCVDTVNSTQFNPQNLTAEQAPNSMDAILKKNAP 1422

RESULT 5
I40229
arginyl endopeptidase - *Porphyromonas gingivalis*
C;Species: *Porphyromonas gingivalis*
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of arginylginsin, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:G927644; PIDN:BAA05484.1; PID:G927645

Query Match 10.7%; Score 266; DB 2; Length 991;
Best Local Similarity 25.9%; Pred. No. 2e-10;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;

QY 30 TIDADGNGNWTTPPGGTSFAGHNSAICASSASYINFEQPNPNLYLTPPELSLNGG 89
Db 595 SNVSCDYNGAIAATISANGKMF---GSAVVENGTATINLTG-----LTNES 637

QY 90 TLEFWCAQDANYASEHYAVASSTG--NDASNANALLEEVLTAKTVTAPEAIRGTRV 147
Db 638 TLTITV---VGYNKEIVIKTINTNGEPNYPQVSN-----LTATT-----QGOKV 679

QY 148 QGTWYQKTVQAGTKYVAFRHFQCTDFFWINLDDV-EIKANGKRAFDTETPESSTHGEA 206
Db 680 TLKWDAPSTKTATN-TARSVDGIRELVLLSVSDAPELLRSQGAIEIVLEAHDVNDGS- 737

QY 207 PAEWTTIDADGGQWCLSSGOLGWLTAHGGTNVASFVN-----GMAINPNYLLISK 261
Db 738 -GYOILLADHDQYGVIPSDTHLPNCVSPALFAPFEYTPENADPSCSTNMIM-- 794

QY 262 DVTGATKVKYKYAVNDGPPGPHYAVMISKTGTNAGDFTVVFEEPTNGINKGARFGLSTE 321
Db 795 DGTASVNI-----PAGTY-----DAI----- 811

QY 322 ADGAKPOS---VMT-----ERTVDPAGTKYVAFRHYNGSDLYNILLDDIQFTMGSP 371
Db 812 ---AAPQANAKIWTAGQPTKEDDYVEAGKY---HFLMKMGSGDGTETLTISEGG- 863

QY 372 TPTDVTYTVVRDGTKIKEGLTETTFEEDGVATGNHCEVVEKYTAGVSPKECVNVTVDPV 431
Db 864 --SDYTVTVVRDGTKIKEGLTETTYRDAAGSAQSHEVCEVVEKYAAGVSPKVCVDYIPDV 921

QY 432 QFNPNQN-LTGSVAVGQKVTIL 450
Db 922 ADVTAQKPYTLTVVGKTIIV 941

RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - *Agrobacterium tumefaciens* (strain C5
C;Species: *Agrobacterium tumefaciens*
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2959
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C59.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:gl7741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3276
A;Map position: linear chromosome

Query Match 5.6%; Score 139.5; DB 2; Length 1052;
Best Local Similarity 20.9%; Pred. No. 0.12;
Matches 123; Conservative 67; Mismatches 159; Indels 239; Gaps 33;

QY 13 TTTLSLSPFENGIPASWKITIDAGDGNWTT-----TPPGGTSFAGHNSAICASSAS 64
Db 52 TVTSGEAILGRHSASEATV--TGDGSKWTTGDLQVGSDTSDPGG--LAGNGT----- 99

```

QY 65 YINFEQPNDNLYVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGN-----116
Db 100 -----LNVTAGSV-----DSTVA--HLGVAGATGSAIVDGKGS 132
QY 117 -----DASNF-----ANALLEEVLTAKTVVTAPEAIRGRVQGTWYQKTVQLPAGTKYVA 166
Db 133 VMTVDNLSLEVSGAGSL---AVTGGGLVDAANIIGTNTGG---NGSVRVSGADSTVK 186
QY 167 FRHFGCTDFWINL-----DDVEIKANG---KRADETETEFESTHGEAPAEWTTIDADGD 218
Db 187 SR-----SD-----LNVGLYNGGSMTEVEAGVAKSRDGYVATYGGST-----SAVTVTGD 231
QY 219 GQWCLCLSSGQLGWLTAHGGT--NVVASFSWNGMALNPDNYLI-----SKDVTGA-TK 268
Db 232 GSSWAMTGFVGY--ASGATGNVTVS---NGAIRATGVTGLDLAGAGSTWTTITGAGSK 286
QY 269 VKYYY-----AVNDGPPG-----DHY-----AVMISKTKGT---293
Db 287 VTAYVDNGTVNGSVDVGFQGGSLSVVNGGSLDAYNLVYVGNALGSSGAVLVSGVSHVS 346
QY 294 -----NAGD-----FTVVFEETPNKNGKARFGLSTEADGAKPQSV 330
Db 347 VDGLMVVGNAGSVBITGASLAAPTILIAEAGSTGVLSIGAGSGQTARSAGA-----401
QY 331 WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ-----FTMGGSPPTDITY 378
Db 402 -VEARAIFAGAGNGSIVFNH---SETGYTLSDISGAGRVAAGVTTLSGNS-----451
QY 379 TVYRDGTKIKEGLTETT-----FEEDG-----VA 402
Db 452 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLNNAISGTSFEXTGDNILL 509
QY 403 TGNHEVCVEVKYTAGVSPKCVNVTVDVQVFNPNVQNLTGSAVGQKVTLL 450
Db 510 TGNSTYSGATAVSAG-----KLSVNGSLASAVSVGSGATVGGTGTI 550

RESULT 7
H98323
Hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98323
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98323
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1341 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:G15160106; GSPDB:GN00170
A:Gene: AGR_L_3085
A:Map position: linear chromosome

Query Match 5.6%; Score 139.5; DB 2; Length 1341;
Best Local Similarity 20.9%; Pred. No. 0.16;
Matches 123; Conservative 67; Mismatches 159; Indels 239; Gaps 33;

```

```

QY 13 TTTLSSEFNGIPASWKTIADGDGNNWTT-----TPPFGTTSFAGHNSAICASSAS 64
Db 341 TVTSGEAIIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDFGG--LAGNGI-----388
QY 65 YINFEQPNDNLYVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGN-----116
Db 389 -----LNVTAGSV-----DSTVA--HLGVAGATGSAIVDGKGS 421
QY 117 -----DASNF-----ANALLEEVLTAKTVVTAPEAIRGRVQGTWYQKTVQLPAGTKYVA 166
Db 422 VMTVDNLSLEVSGAGSL---AVTGGGLVDAANIIGTNTGG---NGSVRVSGADSTVK 475

```

```

QY 167 FRHFGCTDFWINL-----DDVEIKANG---KRADETETEFESTHGEAPAEWTTIDADGD 218
Db 476 SR-----SD-----LNVGLYNGGSMTEVEAGVAKSRDGYVATYGGST-----SAVTVTGD 520
QY 219 GQWCLCLSSGQLGWLTAHGGT--NVVASFSWNGMALNPDNYLI-----SKDVTGA-TK 268
Db 521 GSSWAMTGFVGY--ASGATGNVTVS---NGAIRATGVTGLDLAGAGSTWTTITGAGSK 575
QY 269 VKYYY-----AVNDGPPG-----DHY-----AVMISKTKGT---293
Db 576 VTAYVDNGTVNGSVDVGFQGGSLSVVNGGSLDAYNLVYVGNALGSSGAVLVSGVSHVS 635
QY 294 -----NAGD-----FTVVFEETPNKNGKARFGLSTEADGAKPQSV 330
Db 636 VDGLMVVGNAGSVBITGASLAAPTILIAEAGSTGVLSIGAGSGQTARSAGA-----690
QY 331 WIE-RTVDLPAGTKYVAFRHYNCSDLNYILLDDIQ-----FTMGGSPPTDITY 378
Db 691 -VEARAIFAGAGNGSIVFNH---SETGYTLSDISGAGRVAAGVTTLSGNS-----740
QY 379 TVYRDGTKIKEGLTETT-----FEEDG-----VA 402
Db 741 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVVDGGGTLNNAISGTSFEXTGDNILL 798
QY 403 TGNHEVCVEVKYTAGVSPKCVNVTVDVQVFNPNVQNLTGSAVGQKVTLL 450
Db 799 TGNSTYSGATAVSAG-----KLSVNGSLASAVSVGSGATVGGTGTI 839

RESULT 8
B75622
Hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75622
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaleski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036996; PMID:10567266
A:Accession: B75622
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-631 <WHI>
A:Cross-references: GB:AE001826; NID:G6460827; PIDN:AAF12628.1; PID:G6460924; TIGR:DRB01
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRB0037
A:Map position: megaplasmid
A:Genome: plasmid
A:Note: plasmid Mpi

Query Match 5.3%; Score 132.5; DB 2; Length 691;
Best Local Similarity 21.9%; Pred. No. 0.21;
Matches 111; Conservative 46; Mismatches 178; Indels 171; Gaps 23;

```

```

QY 4 NPENPNPGTTTSLSEFNGIPASWKTIADGDGNNWTTTP-----PPGTSF 51
Db 230 SPNPLPPTGTIT-----PGTGTGPGVSGNGPAESTGPATYTGISTGTPPVVDA 280
QY 52 AGHNSAICASSASYINFEPP-----QNPD-----NLYVTPELS-LPNGGTLTFWVCAQDA 100
Db 281 SONQVAYPKDAETVNPDSVTWTTIIINPNPTAATYELVPLDLSGLPAGVTVTF-----TDA 336
QY 101 N-----YASEHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGRVQGTWYQ 153
Db 337 NGNPLPDDTDGNGREVTAGPNGTA-----TYRVVTVYPTDESAAAAG-----380
QY 154 KTVQLPAGTK-----YVAFRHFGCTDFWNLDDVEIKANGKRADET---195
Db 381 -PIRIPGVGDGNRDGIVDATVTVYNVLLSNLKFNGTNGTALGVSDVPVTRVTPQSVITAV 439

```

```

Qy 196 ----ETTESSTH-----GEAPAEWTTTIDADGGQGWLTCLSSGOLGWLTAHGTT 239
Db 440 VFPMDLFNDGAYDGNALSGSTPIGPVKYATNFTDGG-----VLSPAELALPAE--- 492
Qy 240 NVVASFSWNGMALPNPDNYLISKVGTGATKKYKYVAVNDGFGPDHYAVIMSKTGNAGDFT 299
Db 493 --IAS-----TGAVPVKTEKTV-----YAVVTIPAGQAPGDYM 523
Qy 300 VVFEETPENGINKGARFGLSTEADGAKPQ--SVMTERTVDLPAGT----- 342
Db 524 VT--QTATGSLSGTTKSFNTDKVTVTPSNGSLLIAKRVTTPTTSLNATANGPDVAVS 581
Qy 343 KYVAFRHYNCSDLYIILLDDIQFTMG-----GSPTPTDYTYTV-----YRD-- 383
Db 582 TVTATNNYNTSLYGLVLRDPSNNILGFSFSSNVFGFIKPSSLRATVSGVSGATVLYRTSN 641
Qy 384 -GTKIKEGLTE--TTPEEDGVATGNH 406
Db 642 LNTWAAQPTVDANTTWEVGVDITNNN 667

RESULT 9
A83412
Hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83412
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: A83412
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2468 <STO>
A/Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AA05263.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA1874

```

```

Query Match      5.3%; Score 131.5; DB 2; Length 2468;
Best Local Similarity 22.2%; Pred. No. 1.3;
Matches 120; Conservative 47; Mismatches 184; Indels 189; Gaps 30;

QY      5 PNNPNPCTT-----TJSESPENGIPASWKTIDADGDN----- 38
Db      754 FVINPNSGTTLSGTAEFGSSVTLTDG--NGNPIC--QVTDAGSGNWSFTSTPLADGTVV 809
QY      39 NWITTPPPGCTSPAGHN-----SAICASSASYINPEGPQN----- 74
Db      810 NATATPDAGNWSQGGSTVDGVAPTPTNLNSGSLSGTAEPGTVTLTDGNGNPIAEV 869
QY      75 -----DNYLVTPELSPLNGGTLTFWCAQDANVASEHYAVYASSTGNDASFNANALLEEV 129
Db      870 TADSGSNWTVTPTPPIANGTVN--VVAQDA-----AGN--SSPGASVTVDSQ 913
QY      130 LTAKTVVTAPEALRGTRVQGTWQKTVQLPAGTKYVAFRFGCTDPFWINLDVEIKANG 189
Db      914 APAAPVNVPS--NGTTLTSGT-----AEPGAT-----VTLTD-----GNG 945
QY      190 KRADFTEFPESSTHGEAPAEWTTIDADGCGW-----ICLSSGQIGLWLTAH---GGTVV 241
Db      946 NPTG-----QVTDAGSG--NWSFTPGTFLANGTVVNASTADPTGNTSA 986
QY      242 VASFSWNGMA-----LNPDNLYLSKDVGTATKVKYVAVNDGPPGDHYAVMISKTG--- 292
Db      987 PASTTVDVSAPAAPVNVPSN--GAEISGTAEPGAIVTLTDGSGNPIGQVTDAGSGNWSF 1043
QY      293 ----TNAGDFTVV--FEETPNNGKGARFGLSTEADGAKFPQSVMIERTVDLPAGTKYVAF 347

```

```

Db      1044 TPSPFLADGTVVNATDPAG-NTGGQG-----STTVDIAIPAT-----PTVNLNSGSSLSG- 1094
Qy      348 RHYNCSDLNILLDD-----IOFTMGSGPPTDYYTYVYRDGDKIKEGLTETTFEEDGV 401
Db      1095 ---TAEPGSVTLLTDGNNGPIAEVTDGS---GNWIVT---PSTPIANGTVNVVAQD-- 1143
Qy      402 ATGNHGYCYEVKYTAGVSPKECVN-----VTVPDVQFNPVQLTGSAVG 445
Db      1144 ASGNSSPPATVTVDDSSAPPAPFVNPVSGVVISGTABAGATVTLTDACGNBIGQTADGSG 1203


RESULT 10
C89874
autolysin [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: C89874
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogucuma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kallio, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; PMID:21311952; PMID:11418146
A:Accession: C89874
A>Status: preliminary
A:Molecule type: DNA
X:Residues: 1-1248 <KUR>
A:Cross-references: GB:BA000018; PID:g13700854; PIDN:BAB42150.1; GSFDB:GM00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: atl

```

Query Match 5.2%; Score 128.5; DB 2; Length 1248;
Best local Similarity 22.3%; Pred. No. 0.84;
Matches 105; Conservative 47; Mismatches 165; Indels 153; Gaps 26;

```
QY      3 PNPDPNPQCTTT-----LSESPENGIPASWKIIDADGDG---NNWTTTTPP----- 46
Db      586 PTPPTPKPSTPTNNKLTAVSSLVG---AQINAKNGLFTTVYDQTKPTKEVOKTFA 641
QY      47 -----GGTSF-----AGHNSA-----ICASSASYINFEGPON--PDNYLVTP-----LSPN 87
Db      642 VTKEASLGKNGKPYLVKDYNSPTLLGWVKQGDVLYNNAKSPVNVNQYTVYKPGTKLYSPW 701
QY      88 G-----GTTTFWCAQDANYASEHYAVYASSTGNDASNFAN-----ALLE 127
Db      702 GTYKQEAGAVSGTGNQTFKATKQQQDKS-----IYLFGTVNGKSGWVKAYLAVPAAPKK 757
QY      128 EVLTAKT-----VVTAPAIR-----GTRVQGTWYQKTVQLPAGTKYVAFRHFG 171
Db      758 AVAQPKTAVKAYTVTPQTTQTVSKIAQVKPNNTGIRASVYEKTA--NGAKY-AQRTFY 814
QY      172 CT-----DFFVINLDDVEIKANGKRADPTE--TFESSTHGAPAE 209
Db      815 VTKEAHGNETYVLLNNTSHNIPLGFWNKDLNVQNLGKEVKTTQKYTVKNSNNGLSMPV 874
QY      210 WTTIDADGGQGWLCSSQGLWLTAGHTGNTVVASPSWN-----GMALNPONYLISKDV 263
Db      875 WGP-----KNQVILTGNNI-----AQGTFNATKQVSGVKGVLYGTTINRRTGWNNAKDL 923
QY      264 TGATKVK-----YVYAVNDGPPGHYAVMTLSKTGT-----NAGDFTVVFPEETN 307
Db      924 TAPTAVKPTTSAKDNYTVYVING--NGYVYVTPNSDTAKYSLKAFNEQPFVAVKEQVIN 982
QY      308 G-----INKGAREGLSTEADGAK-----PQSVMIERTVDLPAGTKY 344
Db      983 GQTWYVGKLSNGKLAWIKST--DLAKELIKYNTQGTMTLNQVAIQIAGLQY 1030
```

RESULT 11
AH2515
hypothetical protein alr7304 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120aol
C;Species: Nostoc sp. PCC 7120

RESULT 11
AH2515
hypothetical protein alr7304 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C;Species: Nostoc sp. PCC 7120

A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2515
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:2159285; PMID:11759840
A:Accession: AH2515
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4936 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA078388.1; PID:G17135842; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7304
A:Genome: plasmid

Query Match 5.2%; Score 128; DB 2; Length 4936;
Best Local Similarity 20.4%; Pred. No. 5.5;
Matches 108; Conservative 60; Mismatches 171; Indels 190; Gaps 27;
QY 1 GTPNPNPNGT-----TTLSSEFENGIPASWKTIDAGDGNWTT-----TPPPCG 48
Db 3161 GLSNFN---TPGYQLQVDAIVQDNAGN-----SGDAARTFTTIAAPTGV 3206
QY 49 TSPAGHNSAICASSASYINFEGPQPNLYLV-----TPELSLP-NGG-----TLTF 93
Db 3207 T-----ITQGGSTAVIEG-GNTDSYTLVLRQPTADVTLNTGSGITDKTLTF 3257
QY 94 ----WVCAQ-----DANYASEHYAVASTGNDASNFANALLEEVLTAKTVTAPEA 141
Db 3258 TSANNTPTQITVNAVNDITTEGNTSTISHSISSTDNYSNVLPLDI--AVSITDND 3315
QY 142 IGRTRVQGTWYQKTVLPAGTKYVAFRHPGCTDFF--WINLDDVEIKANGKRA--DETEFE 199
Db 3316 IAGMK-----WNDIGGVKDTGEPGQGWYILD 3345
QY 200 STHGEAP--AEWTTIDAGDGGQCLSSG-----QLGWLTAHGGTNY-----241
Db 3346 SNTNGQLDNGELSTTIDANGNYQ-FNLNRPYVTVAEVQPCWKQTFPGCTNITTADIP 3404
QY 242 ----VASFWSNGMALN--PDNYLSKDVGTGATKYVAVNDGPGDHYAVMSKT 291
Db 3405 LAIPSLDMISPGDSNGIQLNFAANYIKVEDGTAITEV---WVTRGNTSSAVSATLSFT 3461
QY 292 -GTNAG-----DFTVVFEF-----TPNGIN-----310
Db 3462 DGTATGCGGASSVNNDFNNVFTTAFENETSKLISVONALLANPNALKIRNDSKVEGN 3521
QY 311 ----KGARFGLSTEA-----DGAKPQSVMIERTVDLPAG--TKYVAFRHYNCS 353
Db 3522 EYFTIKLNPTGGAVIGNQSIATVITIDDEAPSDITVTPLETPSTTITSADVDSQAIYLI 3581
QY 354 DINYILLDDIQTMGSGPTPTDYTVVYRDGPKIKEGTEITTFEDGVA 402
Db 3582 MLNFWADSRFANIKGN-----DFTSVIIDTGIDLNHPFGADTDNNGIA 3626

RESULT 12
S10789
C:Species: alkaliphilic eubacterium 163-26
C:Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C:Accession: S10789
R:Canussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A:Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an
A:Reference number: S10789; MUID:90336627; PMID:1696201
A:Accession: S10789
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1684 <CAN>

A:Cross-references: EMBL:X53373; NID:G48305; PIDN:CAA37453.1; PID:G48306
Query Match 5.1%; Score 127.5; DB 2; Length 1684;
Best Local Similarity 20.9%; Pred. No. 1.5;
Matches 102; Conservative 53; Mismatches 184; Indels 149; Gaps 27;

QY 23 GIPASWKTIDAGDGNWTTTPPGGTGFAGHNSAICA-SSASYINFEGPQPNLYLVTP 81
Db 204 GLPRDWTNQA--QGNWHT-----HNDINKNDEAAWAWNGSD-----241
QY 82 ELSLNGGTLTFWVCAQD-ANYASEHYAVIASSTGNDASNFANALLEEVLTAKTV-VTAP 139
Db 242 ----WIRADETAGY-----DNCGSEQTMCIGFLPDIKTEVTTGVDLP 280
QY 140 EAIRGT-RVQGTWYQKTVLPAGTKYVAFRHPGCTDFF--WINLDDVEIKANGKRADETE 196
Db 281 PILRNKNDQASGYEDWF-VPAAEFYRQDLNIAPKDYLIKWITSWVEBFGLDPRVDITAK 339
QY 197 TPESSTHGEAPAE-----WTTIDAGDGGQCLSSGQCLWLTA-----HG-----237
Db 340 HVEIERWAEKNEAEVALQTWRENNPDKGANW-----DDNFWMTAEVFGHGLKSEYFD 394
QY 238 -GTNVVAFSW-----NGMALNPONLYISKDVGTGATKYVAVNDGFP 280
Db 395 FGFDVINFEFONANFNLEGLFSRYANSINTDPDNMLSYVSSHDTKL--YSRDD--448
QY 281 GDHYAVMSKTGN---AGDFTVVF-EETPNKNGKARFGLSTEADGAKPQSVW--IE 333
Db 449 -----LIQAGTALLLLPGVQVPGDETARPLGDDG-----SDEQOTRSSMMWANIN 496
QY 334 RTVDLPAGTKYVAFR-----HYNCSDLNYI-----LLDDIQTMGSGPTPT 374
Db 497 QNV-LSHWKLQGFENNHAIAGAHQKLSDSPTFFARTYBESDDIVDEVVATGAQTTA 555
QY 375 DYTIVYRDGPKIKEGLT--ETTPEEDGAVTGNHCVVEYKYTAGVSKPCVNVTVDPVQ 432
Db 556 VTVEGVFEDGTWVRDAYTGDDETTVTK-GTAT-----FTAGTQGIILLIENTAEPVT 604
QY 433 FNPVQNL 440
Db 605 NLPVVSAT 612

RESULT 13
B75489
hypotheical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75489
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: B75489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <WHI>
A:Cross-references: GB:AE001925; GB:AE000513; NID:G6458383; PIDN:AAF10267.1; PID:G64583
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0685
A:Map position: 1

Query Match 5.1%; Score 127; DB 2; Length 713;
Best Local Similarity 21.6%; Pred. No. 0.5;
Matches 129; Conservative 47; Mismatches 214; Indels 208; Gaps 29;
QY 2 TPNPNP-NPNPGTTLSE---SPENGIPASWKTIDAGDGNWTTTTP---PGT-----49
Db 86 TPSTPNPNPSETIETTPPVITVNVNPFPSFTIT---PNDGSKDTTQPDYTNPGQREVK 142

```
QY 50 -----SFAGHNSAICASSASYINFE-----QPONPDNYLVTP-----LSLPGGTL-- 91
Db 143 PCDXNWF-----AYLTNGVNGESYTLTWPTDGTGAVKTPENIRFVLDNNGNQDQ 197
QY 92 -----TFWCAQADANYASEHYAV-----ASSTGNDSANFAN--ALLE 127
Db 138 SEIAAGATTITIDVAINOTVKFQVYDVPTATSDTKFGDPTGTRNDNPNFSDPTLPR 257
QY 128 EVLTAKTV-----TAPEIRGRVQGTWYQKTV-QLPA 160
Db 258 DANNSTVTINRKGWVIGPKADPDGNGPNVTPPAINSPGINIVPTASTQVATVTLPT 317
QY 161 GTKYVAFRHFGCTDFWINLDDVEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGQ 220
Db 318 SGVTVFTN-----TIQNTGNR---IDTFLTQNTFPAGTTVVFKDANGN 360
QY 221 GWLCLSSGQLGWLTAHGGTNVVASFSWNGMALPNPNYLISKDVTGATKYKYVAVNDGFP 280
Db 361 A-----LPTVNGKPVQVQNPENGTV---DIQVITLPAQVTPQL-----SGQP 401
QY 281 GDHYAVMISKGTN---AGDFT--VVEETPNGINKGARFGLSTEADGAKPQSVWIERT 335
Db 402 ---AVTWTTSQNDPTKSDTTKDIIEVKP-----GIAFGDPTPLGGLDPTPVGTPT 451
QY 336 ---VDLPAG---TKYVAFRHNCSDLNYILLDDIQFTMGG-----SPT 372
Db 452 GVPNGPTPLTPGNPQTCTAPITRTYLPMEIANLGSDDAFVVGTAAPVTVLNPDPGTVNPT 511
QY 373 PTDYTYTYVRD---GTKIKEGLTETTFREDGVATGNHEYCEVVKY----- 414
Db 512 FVIVPVVYRYRDNVNGDKLDAGDTL---QGGNTGTIRKPGEEIKLIAVVVDVPCAAQAQTI 567
QY 415 ---TAGVSPEK-----CVNVTVDVPQVQNPVQNLTGSAVGOKVT 449
Db 568 TLNQEKSPITGVSKDPNDITVCGNGKPIVTKVDKARANPGDTLTYTIIGKNTS 625

RESULT 14
A86685
prophage p11 protein 46, tail component [imported] - Lactococcus lactis subsp. lactis (s
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: A86685
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: A86685
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:Cross-references: GB:AE005176; PID:gl12723361; PIDN:AAK04579.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: p1146

Query Match 5.1%; Score 126; DB 2; Length 1441;
Best Local Similarity 22.3%; Pred. No. 1.5;
Matches 98; Conservative 49; Mismatches 133; Indels 160; Gaps 26;

QY 30 TIDADGNNWTTTPPGGTGFGHNSAICASSASYINFEQPONP-----DNYLVTPELSL 85
Db 602 TISSG-----TVPTTGTWTS-----QVPTLVKEQYL----- 628
QY 86 PNGGTLTFWVCAQADANYASEHYAV-YASSTGN-----DANFANALLEEVLTKTV 135
Db 629 ---WTKLWTYTD---NTSETGYSVSYAKOQNGHGDGPPGKDGVGISNTIIEYV----- 677
QY 136 VTAPEAIRGT-RVQGTWYQKTVQLPAGTKYVAFRHF-----GCTDFFWINL----- 180
Db 678 ---GAVSGTSKPTGGWSTTITPTVAG-QYLWTRTWTQYTDGTSEQGVINAMGLTGASG 732
QY 181 -DDVEIKANGKRAADFTETTESSTHG-EAPAE-----WTTIDADG 217
```

```
Db 733 RDGIAGK-DGKIKATAITYQASTNGTTPACTGTWSTVSPSVAKGSFLWTRTIWTYTD--- 788
QY 218 DQGWLCCLSSGQLGWLTAHGGTNVVASFSWNGMALPNPNYLISKDVTGATKYKYVAVN- 276
Db 789 ---NTTETGVAVAYMGIN-----GNGNGINGIAGKDGTTGLKTTIYAVGT 830
QY 277 -----DGFPGDHYAVMISKGTNAGDFTVVFETPNGINKGARGFLSTEA 322
Db 831 SGTTPAGTGMNSQVENVPAQY--LWTKT---VMDYTDKTSYTSVSKFGEK-GDKGDQ 884
QY 323 DGAKPQSVWIERTVDLPAG---TKYVAFRHNCSDL-NXVILLDD-----IQFTWG 368
Db 885 GVQGTQGVGQGGTGPCKAGDKTQYTHIAYANSADGKTNFSTSDSNRTYIGMYVDFNIN 944
QY 369 GSPTPTDYTYTYVR--DGTK 386
Db 945 DSTTPSDYSWTLVKAGDGTQ 964

RESULT 15
C86822
hypothetical protein yqbk [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86822
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1649 <STO>
A:Cross-references: GB:AE005176; PID:gl12724583; PIDN:AAK05677.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yqbk

Query Match 5.1%; Score 126; DB 2; Length 1649;
Best Local Similarity 20.4%; Pred. No. 1.8;
Matches 110; Conservative 56; Mismatches 189; Indels 184; Gaps 28;

QY 46 PGGTSEFAGHNSAICASSASYINFEQ-PQN-----PDNY-----LVTELSLPN 87
Db 65 PSGTTFNG-----NSVSPNRTPLPQNSVEDAIMPNSFGWQQTIANLTLPNQVTAQV 117
QY 88 GGTLTFWVCAQADANYASEHYAVYASSTGNDASNF-----ANALLEEVLTKTVVTAPEA- 141
Db 118 NGTIIGSYPTTDAGVQQAALVAMY-SATNASTSDFLYGGANTLLNSL-LKAVVSNFSAT 175
QY 142 -----IRGTRVQGTWYQKTVQL-----PAGTKYVAFR---HFGCTDFFWINLDDV 183
Db 176 NMTFTLKGHAKSLTWISNPADLLTSSNSQPSGNTYGTLPNSNVYFGVPTIF----RN 231
QY 184 EIKANGKRAADFTETTESSTHGEAPAEWTT-----IDADGQGG 222
Db 232 TVAASGNIVYAQNAPATFNG-----SWITGAPNIYGGTDNSDISGNTNLXIGATGSIAG 287
QY 223 LCLSSGQLGWLTAHGGTNV-----VASFSWNGMALPNPNYLISKDVTG--ATKV 269
Db 288 -NIYGNASAAATISGNTHVTIAQSSSTINSVTGGASGTTISGNTNL---DISCAIASQI 343
QY 270 KYIYAVNDGPPGDHYAVMISKGTNAGDFTVVFETPNG-----INKGARFG-----LS 319
Db 344 TNYGAGIG-----TSNSPVNVNGNVTYVNSTNGGARVQLYQGGTYVYGNISSIY 394
QY 320 TEADGAKPQSVWIERTVDL-----PA-----GTKYVAFRHNCSDLN 359
Db 395 NTLUGA---GGWTGATSNINGAGGPASTFNGSGFQGNIGTSGAGNVISNSYTSSTF--- 448
QY 360 LDDIQFTWGGSPPTDYTYTYVRDGTGKIKELTETTFEEDGVATGNHEYCEVVKYTAG-- 417
```

Db 449 TGOALFTGCGNAGTASYAQA-----TNSTTAAQGIILYANITNVIKSAFTTGTA 496
QY 418 -----VSPKE-----CVNVTVDVPQFNPVQNLTGSAVGOKVT 449
Db 497 GAVYGIVGGNGHDSLKISPSQWGLGSGTGLDSAVGT-DAKAYGQIPSTTVVSNQKIT 554

Search completed: May 18, 2004, 11:47:49
Job time : 13.906 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 8.3342 Seconds
(without alignments)
2848.981 Million cell updates/sec

Title: US-08-570-311-18

Perfect score: 2480

Sequence: 1 GTNPNNPNNPGTTTLESF.....QNLTSVAVGQVTLKWDAPN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 2480 | 100.0 | 2628 | 1 | HGA2_PORGI |
| 2 | 2441 | 98.4 | 2164 | 1 | HGA2_PORGI |
| 3 | 1453 | 58.6 | 989 | 1 | PRTH_PORGI |
| 4 | 266 | 10.7 | 991 | 1 | CPGI_PORGI |
| 5 | 131.5 | 5.3 | 1256 | 1 | ATL_STAAR |
| 6 | 126 | 5.1 | 721 | 1 | OGP_MOUSE |
| 7 | 121.5 | 4.9 | 465 | 1 | SLAP_LACBR |
| 8 | 121 | 4.9 | 1090 | 1 | GUXB_CELFI |
| 9 | 118.5 | 4.8 | 1385 | 1 | C5AA_BACUD |
| 10 | 114 | 4.6 | 872 | 1 | GUXA_CELFI |
| 11 | 113 | 4.6 | 2660 | 1 | YEEJ_ECO57 |
| 12 | 112.5 | 4.5 | 1034 | 1 | BGAL_BACME |
| 13 | 112.5 | 4.5 | 1260 | 1 | ALSI_CANAL |
| 14 | 110.5 | 4.5 | 1122 | 1 | ADPI_MYCGA |
| 15 | 109.5 | 4.4 | 331 | 1 | PME_ASPAC |
| 16 | 109.5 | 4.4 | 607 | 1 | YK28_ARCFU |
| 17 | 108.5 | 4.4 | 1953 | 1 | BIGA_SALTY |
| 18 | 108 | 4.4 | 699 | 1 | CHII_BACCI |
| 19 | 108 | 4.4 | 987 | 1 | YD94_METJA |
| 20 | 108 | 4.4 | 1289 | 1 | C5AB_BACUD |
| 21 | 107.5 | 4.3 | 551 | 1 | AMVB_THETU |
| 22 | 107.5 | 4.3 | 1200 | 1 | HYAL_STRPU |
| 23 | 107 | 4.3 | 1656 | 1 | OMPB_RICJA |
| 24 | 107 | 4.3 | 2249 | 1 | OMPA_RICRI |
| 25 | 106.5 | 4.3 | 1122 | 1 | ADP2_MYCGA |
| 26 | 106.5 | 4.3 | 2124 | 1 | YEEU_RAT |
| 27 | 106.5 | 4.3 | 2358 | 1 | YEEU_ECOLI |
| 28 | 106 | 4.3 | 1045 | 1 | PRTS_SERMA |
| 29 | 106 | 4.3 | 1300 | 1 | 120K_RICRI |
| 30 | 106 | 4.3 | 1654 | 1 | OMPB_RICRI |
| 31 | 105.5 | 4.3 | 1220 | 1 | C5AC_BACTU |
| 32 | 105.5 | 4.3 | 1534 | 1 | YFAS_ECO57 |
| 33 | 105 | 4.2 | 827 | 1 | XANP_XANS2 |

| | | | | | |
|----|-------|-----|------|---|------------|
| 34 | 105 | 4.2 | 1045 | 1 | GUNB_CELFI |
| 35 | 105 | 4.2 | 2812 | 1 | ZAN_HUMAN |
| 36 | 103.5 | 4.2 | 837 | 1 | XINZ_CLOTM |
| 37 | 103.5 | 4.2 | 1655 | 1 | OMPB_RICCN |
| 38 | 103.5 | 4.2 | 1861 | 1 | APU_THETU |
| 39 | 103 | 4.2 | 1157 | 1 | C8AA_BACUK |
| 40 | 102.5 | 4.1 | 282 | 1 | PRTA_ASPNG |
| 41 | 102.5 | 4.1 | 1045 | 1 | PRTT_SERMA |
| 42 | 102 | 4.1 | 435 | 1 | AM3D_ORISA |
| 43 | 102 | 4.1 | 941 | 1 | GUN_BACS6 |
| 44 | 101.5 | 4.1 | 1462 | 1 | PTP6_DROME |
| 45 | 101.5 | 4.1 | 2132 | 1 | PGCA_MOUSE |

P26225 cellulomona
Q9Y493 homo sapien
P10478 clostridium
Q9Kka3 r outer mem
P38536 t amylopull
Q45704 bacillus th
P24665 aspergillus
P29805 serratia ma
P27933 oryza sativ
P19424 bacillus sp
P16620 drosophila
Q61282 mus musculu

ALIGNMENTS

RESULT 1
HGA2_PORGI
ID_HGA2_PORGI STANDARD; PRT; 2628 AA.
AC Q51845;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin A precursor.
GN HGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97047672; PubMed=8926061;
RA Han N., Whitlock J., Proguiske-Fox A.;
RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
contains four large, contiguous, direct repeats.";
RL Infect. Immun. 64:4000-4007(1996)
CC -!- FUNCTION: Agglutinates erythrocytes.
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U41807; AAB17128.1; -;
DR PIR; T28651; T28651.
KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 995 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 100.0%; Score 2480; DB 1; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.6e-163;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTNPNNPNNPGTTTLESFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICA 60
Db 950 GTNPNNPNNPGTTTLESFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICA 1009
QY 61 SSASYINFGPQNDNYLVTPELSLNGGTLTFWVCAQDANYASEHYVYASSTGNDASN 120
Db 1010 SSASYINFGPQNDNYLVTPELSLNGGTLTFWVCAQDANYASEHYVYASSTGNDASN 1069

QY 121 FANALLEEVLTAKTIVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDREWINL 180
DB 1070 FANALLEEVLTAKTIVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDREWINL 1129
QY 181 DVEIKANGKRAADFTTFFSSSTHGEAPAEWTTIDADGGQGLWLTAGHGTN 240
DB 1130 DVEIKANGKRAADFTTFFSSSTHGEAPAEWTTIDADGGQGLWLTAGHGTN 1189
QY 241 VVASFSWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTV 300
DB 1190 VVASFSWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTV 1249
QY 301 VFEETPNKGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDNLYILL 360
DB 1250 VFEETPNKGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDNLYILL 1309
QY 361 DDIOFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
DB 1310 DDIOFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 1369
QY 421 KECVNVTVDPVQFNPVONLTGSAVGOKVTLKWDAPN 456
DB 1370 KECVNVTVDPVQFNPVONLTGSAVGOKVTLKWDAPN 1405

RESULT 2
HGAL PORGI
ID AC P5915; STANDARD; PRT; 2164 AA.
DT 10-OCT-2003 (Rel. 42, Created)
DT 15-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hemagglutinin A precursor.
GN HAGA OR PG1837.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=22829867; PubMed=12949112;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA Dewhirst F.E., Fraser C.M.;
RT "Complete genome sequence of the oral pathogenic bacterium
Porphyromonas gingivalis strain W83.";
RL J. Bacteriol. 185:5591-5601(2003).
CC -!- FUNCTION: Agglutinates erythrocytes (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB017178; AAQ66831.1; ALT_INIT.
DR TIGR; PG1837;
KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 2164 HEMAGGLUTININ A.
FT DOMAIN 26 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 991 PEPTIDASE C25-LIKE 2.
FT DOMAIN 992 1443 PEPTIDASE C25-LIKE 3.
SQ SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;
Query Match 98.4%; Score 2441; DB 1; Length 2164;

Best Local Similarity 98.9%; Pred. No. 6.2e-161;
Matches 449; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 3 PNPENPNPGTTLTSESFENGIPASWKITIDADGNNWTTTTPPGGTSPAGHNSAICASS 62
DB 496 PNGTENPNPGTTLTSESFENGIPASWKITIDADGNNWTTTTPPGGTSPAGHNSAICASS 555
QY 63 ASYINFEQPNPDNLYLTPVPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFA 122
DB 556 ASYINFEQPNPDNLYLTPVPELSPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFA 615
QY 123 NALLEBEVLTAKTIVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDREWINLDD 182
DB 616 NALLEBEVLTAKTIVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDREWINLDD 675
QY 183 VEIKANGKRAADFTTFFSSSTHGEAPAEWTTIDADGGQGLWLTAGHGTN 242
DB 676 VEIKANGKRAADFTTFFSSSTHGEAPAEWTTIDADGGQGLWLTAGHGTN 735
QY 243 ASFSWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTV 302
DB 736 ASFSWNGMALNPNDNLYLSKDVGTGATKVKYVAVNDGPPGDHYAVMISKTGTNAGDFTV 795
QY 303 EETPNKGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDNLYILLDD 362
DB 796 EETPNKGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDNLYILLDD 855
QY 363 IOFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKE 422
DB 856 IOFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPKE 915
QY 423 CUNVTVDVQFNPVONLTGSAVGOKVTLKWDAPN 456
DB 916 CUNVTVDVQFNPVONLTGSAVGOKVTLKWDAPN 949

RESULT 3
PRTH PORGI
ID AC P46071; STANDARD; PRT; 989 AA.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protease prth (EC 3.4.22.-).
GN PRTH.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=95012612; PubMed=7927685;
RA Fletcher H.M., Schenkeln H.A., Macrina F.L.;
RT "Cloning and characterization of a new protease gene (prth) from
Porphyromonas gingivalis.";
RL Infect. Immun. 62:4279-4286(1994).
RN [2]
RP ERRATUM.
RA Fletcher H.M., Schenkeln H.A., Macrina F.L.;
RL Infect. Immun. 62:5707-5707(1994).
CC -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
CC P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
CC IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
CC AND IS A VIRULENCE FACTOR.
CC -!- SUBCELLULAR LOCATION: In membrane vesicles.
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

QY 432 QPNPQVQ-LTGSVAGOKVTL 450
 Db 922 ADVTAQKPVTLTVGKTIIV 941
 RESULT 5
 ATL_STAATU STANDARD; PRT; 1256 AA.
 AC P52081;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bifunctional autolysin precursor [includes: N-acetylmuramoyl-L-alanine
 DE amidease (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
 DE acetylglucosamidase (EC 3.2.1.96)]
 GN ATL.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
 RC STRAIN=RN450;
 RA OSHIDA T., Sugai M., Komatsuzawa H., Hong Y.-M., Suganaka H.,
 RA Tomasz A.;
 RT "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
 RT alanine amidease domain and an endo-beta-N-acetylglucosaminidase
 RT domain: cloning, sequence analysis, and characterization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIC 8325-4;
 RA Foster S.J.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN
 CC HIGH-MANNOSE GLYCOPROTEINS AND GLYCOPROTEINS CONTAINING THE
 CC -[(MAN)5(GLCNAc)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
 CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
 CC OLIGOSACCHARIDE IS RELEASED INTACT.
 CC -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
 CC residues and L-amino acid residues in certain bacterial cell-wall
 CC glycopeptides
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
 CC unit in high-mannose glycopeptides and glycoproteins containing
 CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
 CC residue remains attached to the protein; the rest of the
 CC oligosaccharide is released intact.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
 CC EXTRACELLULAR LYtic ENZYMES.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
 CC ACETYL-MURAMOYL-L-ALANINE AMIDASE FAMILY 2.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
 CC GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D17366; BAA04185.1; -
 DR EMBL; L14199; AAA99982.1; -
 DR InterPro; IPR002502; Amidase_2.
 DR InterPro; IPR002901; Amidase_4.
 DR Pfam; PF01510; Amidase_2; 1.
 DR Pfam; PF01832; Amidase_4; 1.
 DR SMART; SM00644; Ami_2; 1.
 DR SMART; SM00047; LY22; 1.
 KW Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
 FT SIGNAL 1 29 POTENTIAL.

FT CHAIN 30 1256 BIFUNCTIONAL AUTOLYSIN.
 FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
 FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
 FT REPEAT 425 589 1.
 FT REPEAT 596 758 2.
 FT REPEAT 770 932 3.
 SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;
 Query Match 5.3%; Score 131.5; DB 1; Length 1256;
 Best Local Similarity 22.3%; Pred. No. 0.24;
 Matches 105; Conservative 47; Mismatches 165; Indels 153; Gaps 26;
 QY 3 PNPENPNPPTTT-----LSESEFENGIPASWKTIIDAGDG-----NNWTTTPPP----- 46
 Db 594 PTPPTPEPTPTTNKTLVSLNGV-----AQINAKNNGLFTTVYDKTGKPTKEVQKTF 649
 QY 47 -----GGTSF-----AGHNSA-----ICASSASYINFEGPQN-PDNYLVTPR-----LSLPN 87
 Db 650 VTKEASLGGNKEFLVKDYNPTLIGWVKQGDVYNNAKSPVNMQTYTVKPGTKLYSVEW 709
 QY 88 G-----GTLTFVCAQDANYASEHYAVYASSTGNDASNFAN-----ALLE 127
 Db 710 GTYKQEAGAVSGTGNQTEKATKQQIDKS-----IYFGTVNGKSGVSKAYLAVPAAPKK 765
 QY 128 EVLTAKT-----VVTAPPAIR-----GTFVQGTWYQKTVQLPAGTKYVAFRHFG 171
 Db 766 AVAQPKTAVKATVTPKPTQTVSKIAOVKNNTGIRASVVEKTAK--NGAKY-ADRTFY 822
 QY 172 CT-----DFFWINLDDVEIKANGKRADFTB--TFEGSTHCEAPAE 209
 Db 823 VTKEAHGNETVYLLNNTSHNIPLGWENVKDLNQLGKEVKTQKYTVNKSNGLSMVP 882
 QY 210 WTTTADGGGQWCLSSGQGLWLTAHGNTNVVASFSWN-----GMALNPNDNYLISKDV 263
 Db 883 WGT-----KNQVILTCNNI-----AQGTENATKQSVGKDVLYLGTINNRTGWNAKDL 931
 QY 264 TGATVKV-----YYAVNDGFGPDHYAVMISKTGT-----NAGDFTVYFEETPN 307
 Db 932 TAPTAKVETTSAAKDNYTVYVYKNG-NGYIYVTPNSDTAKYSLKAFNEQPPFAVVKEQVN 990
 QY 308 G-----INRGGARFGLSTEADGAK-----PQSVMIERTVDLPAGTKY 344
 Db 991 GQWYVYGLSKGLAWIKST--DLAKELIKYNTQGTMLNQAQIAGLQY 1038
 RESULT 6
 OGP_MOUSE STANDARD; PRT; 721 AA.
 ID OGP_MOUSE
 AC Q62010;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
 DE (Oviductin) [Estrogen-dependent oviduct protein].
 GN OVGP1 OR OGP OR CHIF5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Oviduct;
 RX MEDLINE=96115001; PubMed=7492680;
 RA Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,
 RA Araki Y.;
 RT "Molecular cloning and characterization of a mouse oviduct-specific
 RT glycoprotein."
 RL Biol. Reprod. 53:285-294 (1995).
 CC -!- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
 CC in the fertilization process and/or early embryonic development.
 CC -!- SUBCELLULAR LOCATION: Secretory granules.
 CC -!- TISSUE SPECIFICITY: Epithelial cells of the oviduct.
 CC -!- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

DR EMBL; D32137; BAA06863.1; --
 DR MGD; MGI:106661; Ovpl.
 DR InterPro; IPR001223; Glyco_hydro_18.
 DR InterPro; IPR001579; Glyco_hydro_18AS.
 DR Pfam; PF00704; Glyco_hydro_18; 1.
 DR ProDom; PD000471; Glyco_hydro_18; 1.
 DR SMART; SM00636; Glyco 18; 1.
 DR PROSITE; PS01095; CHITINASE 18; FALSE NEG.
 DR GlycoProtein; Fertilization; Repeat; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 721
 FT DOMAIN 486 632
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 721 AA; 78807 MW; 37246C8F01665652 CRC64;

Query Match 5.1%; Score 126; DB 1; Length 721;
 Best Local Similarity 18.9%; Pred. No. 0.28;
 Matches 103; Conservative 55; Mismatches 185; Indels 202; Gaps 22;

QY 51 FAGHNSAICA-----SSASVINF----- 68
 Db 222 FTGHNSPLFLSPEDSKSSAYAMNWRKLGTPADKLINGFPYGRNLYLKESKNGLOTAS 281

QY 69 EGPONPNYLVTPELSPLNGGTLTFWCAQDANYASEHYAVYASSTGNDASNFANALLE 128
 Db 282 MGPASPKY-----TKQAGFLAYEVCSPVQRAKHWIDYQ----- 317

QY 129 VLTAKTVVTVAPEAIRGRVQGTQKTVQVLPAGTKYVAFRHFCTDFFWINLDDV----- 183
 Db 318 -----YVPYAFKGEWLG--YDDTISFSYKAMYKREHFGAMVWTLMDDDVRGTFC 367

QY 184 -----EIKANGKRAD-----FTETFEEST----- 202
 Db 368 GNGPFPPLVHILNELLVQTESNSTPLQPFWFTSSVNASGPGSENTALTEVLTDTIKILPP 427

QY 203 HGEAPA-----EWTIDADGGQGLCLSSQGLWHTAHGCT---NVVAS----- 244
 Db 428 GGEAMTEVRRYENMTVPSDGS-----VTPGGTASPRKHAVTPENNTMAAEAKTMT 481

QY 245 ----FSNMGMALPNPNYLI SKDVTGATKV-----KYVAVNDGPGDHYAVM-ISKGTGNA 295
 Db 482 LDFFSKTTTGVSKTITGSKTTTGVSKTITGVSATAGISKTIPEISKATAGSKITTVG 541

QY 296 GDETVVFEETPENGKNGARFGLSTEDAGAKPOSWIER-TVLPAGTKYVAFRHNCS 354
 Db 542 SKTTTGSKITITGSK--TTTGISKITGSKITGSKITGSKITGSKITGSKITGSKITG 598

QY 355 LNVILLDDIQTMGSPPTDVTYVVRDGTIKIKEGLTEIT-----FEEDGVAT 403
 Db 599 -----QTTTGISKITITDISKTT--TGISKITPGISKITPGMTVIVQQAEEAETAT 648

QY 404 GNHEYCVVEK-----YTAGVSPKE-----CVNVTVDFVQFNPVONLTGSAVGOK 447
 Db 649 MDHQSVPTPTMTDTLFLYKWTMTSEKTSRKTWVLEKATVSPREMSATFN-----GQS 702

QY 448 VTLKW 452
 Db 703 KTLKW 707

RESULT 7

SLAP LACBR
 ID SLAP LACBR STANDARD; PRT; 465 AA.
 AC Q05044;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE S-layer protein precursor (Surface layer protein).
 OS Lactobacillus brevis.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1580;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 9287 / DSM 20556;
 RX MEDLINE=93054358; PubMed=1429463;
 RA Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
 RT "S-layer protein gene of Lactobacillus brevis: cloning by polymerase
 chain reaction and determination of the nucleotide sequence.";
 RL J. Bacteriol. 174:7419-7427(1992).
 CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
 of proteins which coat the surface of bacteria.
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
 S-layer with tetragonal symmetry.
 CC -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

EMBL; Z14250; CAA78618.1; --
 DR EMBL; Z14250; CAA78618.1; --
 DR PIR; A47023; A47023.
 KW Signal; Glycoprotein; Cell wall; S-layer.
 FT SIGNAL 1 30
 FT CHAIN 31 465 S-LAYER PROTEIN.
 SQ SEQUENCE 465 AA; 48159 MW; 2BE2403932B65A2C CRC64;

Query Match 4.9%; Score 121.5; DB 1; Length 465;
 Best Local Similarity 19.8%; Pred. No. 0.32;
 Matches 99; Conservative 48; Mismatches 201; Indels 153; Gaps 20;

QY 13 TTILSSSFEN-GTPASWKIIDA-----DGDGNNWTTTPPGGTGFAGHNSAICASSAYIN 67
 Db 26 TTASAKSYATAGAYSTLKTDAATRNVEATGTALYTKP--GTV---KGAKVASKATMAK 80

QY 68 FEGPQPNLVLTPELSPLNGGTLTFWCAQDANYASEHYAVYASSTGNDASNFANALLE 127
 Db 81 LASSKKSADYFRAYGVKTNRGSVYRVVTMDGKYGVYVG-----GKSDTAPAGI-- 132

QY 128 EVLTAKTVVTVAPEAIRGRVQGTQKTVQ-----LPAGTKYVAFRHFCTDFFWINLDD 182
 Db 133 -----KSAETTTKADMPARTTGFYLTDSKTLWTAPKYQYKASK----- 173

QY 183 VEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGG-QGWLCCLSSQGLWHTAHGNTV 241
 Db 174 VSLYGVAKDTKFTVDQAATKTRGSLYHYHTATNGSGISGWIYAGK---GFSSTATGTQV 230

QY 242 VASFSNMGMALPNPNYLI SKDVTGATKVYVAVNDGPGDHYAVMISKGTNAGDFTVV 301
 Db 231 LGGLSTDK-----SVTATNDSVKIVRTTD-----GTQVGSNTWV 266

QY 302 FEETPNGKNGARFGLSTEDAGAKPOSWIERTVLPAGTKYV-----AFRHNCS 353
 Db 267 -----TSTDGTGAGSKVSKADQDAQTALAYINANKPSGVTVTPNAA 308

QY 354 DLNY-----ILLDDIQTMGSPPTDYT-----YTVY 381
 Db 309 DATYGNVTYATVSOATSKVALKVSGTPTVTTALTATADANDKVAANDTTANGSVAGSTV 368
 QY 382 RDGTIKIKEGLTETTFE-----EDGVTGNHEYCVVE-----KYTAGVSPK 421

Db 369 AAGTKIAQLTDLTGKGVVLTADTDLEATGTGTTTYSDLGKAYHYTYTNKQSA 422

Qy 422 ECVNVTVDFVQFNPVNQLTGS 442

Db 429 ASSNAS---TQFG--SNVTGT 444

RESULT 8

GUXB CELFI STANDARD; PRT; 1090 AA.

ID AC P50699;

DT DT 01-OCT-1996 (Rel. 34, Created)

DT DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE DE Exoglucanase B precursor (BC 3.2.1.91) [Exocellobiohydrolase B]

DE DE (1,4-beta-cellobiohydrolase B) (CBP120).

GN CBHB OR CENE.

OS Cellulomonas fimi

OC OC Actinobacteria; Actinobacteridae; Actinomycetales;

OC OC Micrococccaceae; Cellulomonadaceae; Cellulomonas.

OX NCBI_TaxID=1708;

RN [1]

RN SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.

RC STRAIN=ATCC 484;

RX MEDLINE=96003898; PubMed=7575482;

RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

RA "Cellulobiohydrolase B, a second exo-cellulobiohydrolase from the

RT RT cellulolytic bacterium Cellulomonas fimi.";

RL RL Biochem. J. 311:67-74(1995).

RN [2]

RN SEQUENCE OF 54-75.

RP MEDLINE=93209933; PubMed=8458833;

RP Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

RA RA "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase

RT RT D (CenD), a family A beta-1,4-glucanase.";

RL RL J. Bacteriol. 175:1910-1918(1993).

RN [3]

RN SEQUENCE OF 54-78.

RP MEDLINE=94197708; PubMed=8147863;

RP Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,

RA RA Warren R.A.J., Miller R.C. Jr.;

RA "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi

RT RT CenE, a member of a new family of beta-1,4-glucanases";

RL RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).

CC -! FUNCTION: Hydrolyzes cellobiose to a mixture of cellobiotetraose,

CC cellobiotriose and cellobiose, with only a trace of glucose. It

CC hydrolyzed cellobiotetraose to cellobiotriose and cellobiose, and

CC cellobiotetraose to cellobiose, but it did not hydrolyze cellobiotriose.

CC Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds

CC with inversion of anomeric configuration.

CC -! CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages

CC in cellulose and cellobiotetraose, releasing cellobiose from the non-

CC reducing ends of the chains.

CC -! SIMILARITY: Contains 3 fibronectin type III domains.

CC -! SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)

CC domain.

CC -! SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL

CC HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to licensese@isb-sib.ch).

CC -----

CC EMBL; L38827; AAC00822.1; --

CC PIR; S59077; S59077.

CC HSP; P07986; LEXG.

CC InterPro; IPR001919; Bac_celose-bind.

CC InterPro; IPR008965; Cellul_bind.


```

DR PRINTS; PR00733; GLHYDLASE6.
DR PRODOM; PD003733; Glyco_hydro_6; 2.
DR SMART; SM00637; CBD II; 1.
DR DR SMART; SM00060; FN3; 3.
DR PROSITE; PS00561; CBD BACTERIAL; 1.
DR PROSITE; PS00655; GLYCOSYL HYDROL_F6 1; 1.
DR PROSITE; PS00656; GLYCOSYL HYDROL_F6 2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 40
FT CHAIN 41 872
FT DOMAIN 41 872 EXOGLUCANASE A.
FT DOMAIN 41 477 CATALYTIC.
FT DOMAIN 478 563 FIBRONECTIN TYPE-III 1.
FT DOMAIN 573 664 FIBRONECTIN TYPE-III 2.
FT DOMAIN 673 768 FIBRONECTIN TYPE-III 3.
FT DOMAIN 769 872 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 188 188 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 410 410 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 140 202 BY SIMILARITY.
FT DISULFID 374 428 BY SIMILARITY.
FT DISULFID 770 869 BY SIMILARITY.
SQ SEQUENCE 872 AA; 89300 MW; 7883B407F995533B CRC64;

Query Match
Best Local Similarity 4.6%; Score 114; DB 1; Length 872;
Matches 92; Conservative 29; Mismatches 146; Indels 112; Gaps 19;

QY 12 GTTTLSEFENGIPASWKTIADGDGNNMTTTPPGGTGFAGHNSAICASSASYINFEGP 71
Db 491 GTTTAT-----SVPLSS---TASTDNVAVTGYDVRGTTVLGTTAA----- 528
QY 72 QNPNDVNLVTPSLPNGGTLTFWVCAQDA--NYASEHYAVYASSTGNDASNFANALLEV 129
Db 529 ----TSVTVT---GLTPATAYSTFVRAKDAGNVSAASAAAAATQSGTVDTTAPSPVAG 582
QY 130 LFAKTVVTAP-----EALRGTRVQGTW---YOKTVOLPAGTKV 165
Db 583 LTAGTTTTTVPLSWTASTDNAGGSGVAGYVLRGTIVVGTGTTATSYVT-GLTAGTIV- 640
QY 166 AFRHFGCTOFFMINLDVIEIKANGRADTFE-----SSTHGEA 206
Db 641 -----SFSVRADKAGNTSAASAASVATQTGTVVDTTAPSVPTGLTAGTITTSV 691
QY 207 PAEWITIDAGGQGLCLSSQLGHLTAHGGTNV--VASFWNGMALNPDN-----YLIS 260
Db 692 PLTWI---ASTDNAG---GSGVAGYEVNGTRVATVTSTYVTVGLAADAAYSFTVKA 744
QY 261 KDVTG-----ATKVKYYAVNDGPPGDHYAVMISKTGNTAGDFTVWFEEFPGINK-G 312
Db 745 KDVAGNVSAASAASVARTQAATSGG-----CTVKYSASSWNTG-FTGIVVEKNGTAALN 798
QY 313 GARFGLSTEADGAKPQSVW 331
Db 799 GWTGLFSP-ADGQKVSQGW 816

RESULT 11
YEEJ_ECO57
AC YEEJ_ECO57 STANDARD; PRT; 2660 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeeJ.
GN Z3135 OR_ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN 1
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

```

```

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Okusubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005423; AAG57041.1; -
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR Pfam; PF02369; Big_1; PKD.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00089; PKD; 8.
DR DOMAIN 738 834
DR DOMAIN 840 929
DR DOMAIN 931 1033
DR DOMAIN 1042 1132
DR DOMAIN 1134 1236
DR DOMAIN 1245 1335
DR DOMAIN 1337 1439
DR DOMAIN 1448 1539
DR DOMAIN 1548 1652
DR DOMAIN 1653 1750
DR DOMAIN 1751 1855
DR DOMAIN 1856 1957
DR DOMAIN 1963 2056
DR DOMAIN 2065 2156
DR DOMAIN 2157 2252
DR DOMAIN 2254 2355
DR DOMAIN 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match
Best Local Similarity 4.6%; Score 113; DB 1; Length 2660;
Matches 104; Conservative 58; Mismatches 199; Indels 106; Gaps 24;

QY 13 TTTLSEFSE---NGIPASWKTIADGDGNNMTTTPPGGTGFAGHNSAICASSASYINF 69
Db 1470 TATVKDQFDNEVNNLPVTFSTA-----SSGLTLPGESNTNESCIAQTLAGVAF----- 1519
QY 70 GPQNPNDVNLVTPSLPNGG-----TLTF---WVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1520 GEQ-----TVTASLANNGASDNKTVHFGTGTAAAKIIBLTVPDPSIIAGTQPNSGSG- 1571

```


| | | | |
|----|--|--|-----------------------------------|
| | ----- | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch) | ----- |
| CC | EMBL; L25902; AAC41649.2; -- | | |
| DR | InterPro; IPR008440; Candida ALS. | | |
| DR | Pfam; PF05792; Candida ALS; I. | | |
| KW | Cell adhesion; Glycoprotein; Repeat; Signal. | | |
| FT | SIGNAL | 1 | POTENTIAL. |
| FT | CHAIN | 18 | AGGLUTININ-LIKE PROTEIN 1. |
| FT | DOMAIN | 433 | 10 X 36 AA TANDEM REPEATS. |
| FT | REPEAT | 433 | 1-1. |
| FT | REPEAT | 469 | 1-2. |
| FT | REPEAT | 505 | 1-3. |
| FT | REPEAT | 541 | 1-4. |
| FT | REPEAT | 577 | 1-5. |
| FT | REPEAT | 613 | 1-6. |
| FT | REPEAT | 649 | 1-7. |
| FT | REPEAT | 685 | 1-8. |
| FT | REPEAT | 721 | 1-9. |
| FT | REPEAT | 757 | 1-10. |
| FT | DOMAIN | 983 | 2 X 26 AA APPROXIMATE REPEATS. |
| FT | REPEAT | 983 | 2-1. |
| FT | REPEAT | 1092 | 2-2. |
| FT | DOMAIN | 399 | POLY-THR. |
| FT | DOMAIN | 408 | POLY-THR. |
| FT | DOMAIN | 450 | POLY-THR. |
| FT | DOMAIN | 486 | POLY-THR. |
| FT | DOMAIN | 522 | POLY-THR. |
| FT | DOMAIN | 558 | POLY-THR. |
| FT | DOMAIN | 594 | POLY-THR. |
| FT | DOMAIN | 630 | POLY-THR. |
| FT | DOMAIN | 666 | POLY-THR. |
| FT | DOMAIN | 702 | POLY-THR. |
| FT | DOMAIN | 738 | POLY-THR. |
| FT | DOMAIN | 774 | POLY-THR. |
| FT | DOMAIN | 874 | POLY-SER. |
| FT | CARBOHYD | 471 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 579 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 615 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 687 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 723 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 820 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 886 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 918 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 973 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 1045 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 1068 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 1260 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| Q | SEQUENCE | 1260 AA; 132641 MW; 763DI063A2354C24 | CRC64; |

[illegible]

```

Db      566 -----PGGTDSVIREPPNHTVTTEYW-----SQSYA 593
Qy      200 SSTHGEPAAEWTTIADAGDGGWCLSSGQLGWTAHGTTNVASFSWNGMALNPONYLI 259
Db      594 TTTTWTAPP-----GGGDTV-----IIREPPNHTV 618
Qy      260 SKDVTGATVKYKYAVANDPFGDHYAVMISKTGNAGDFTVFEETN 307
Db      619 -----TTTEYW-----SQSFATTTVTGPPSGTDTVIIREPPN 651

RESULT 14
ADP1 MYCGA
ID ADP1 MYCGA STANDARD; PRT; 1122 AA.
AC Q49379; Q49437; Q53351;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN GAPA OR MGCI.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=2096;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=S6;
RX MEDLINE=96201559; PubMed=8613358;
RA Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
RT "Cloning and characterization of a putative cytoadhesin gene (mgci)
RN from Mycoplasma gallisepticum.";
RL Infect. Immun. 64:1541-1547(1996).
[2]
RP SEQUENCE OF 1-12 FROM N.A.
RC STRAIN=S6;
RA Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RN Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 378-570 FROM N.A.
RC STRAIN=S6;
RX MEDLINE=93371270; PubMed=8363503;
RA Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
RT "Identification of the putative cytoadhesin gene of Mycoplasma
RN gallisepticum and its use as a DNA probe.";
RL Avian Dis. 37:380-388(1993).
[4]
RP SEQUENCE OF 159-1122 FROM N.A.
RC STRAIN=S6;
RA Goh M.S., Geary S.J.;
RN Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Could be involved in cytoadherence.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the adhesin P1 family.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
RN frameshift in position 159 to 213.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U34842; AB02987.1; -
DR EMBL; U44804; AAC83385.1; ALT_FRAME.
DR F01; T18346; T18346.
FT Cytoadherence; Signal; Transmembrane.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1122 ADHESIN P1.
FT TRANSMEM 997 1021 POTENTIAL.
FT DOMAIN 1002 1009 POLY-ILE.
FT CONFLICT 313 314 DM -> IW (IN REF. 4).
FT CONFLICT 382 382 Y -> F (IN REF. 3)
CC -----

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U34842; AAC02987.1; -;
 DR EMBL; U44804; AAC83385.1; ALT_FRAME.
 DR FIR; T18346; T18346.
 DR Cytoadherence; Signal; Transmembrane.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1122 ADHESIN PL.
 FT TRANSMEM 997 1021 POTENTIAL.
 FT DOMAIN 1002 1009 POLY-TILE.
 FT CONFLICT 313 314 DM -> IW (IN REF. 4).
 FT CONFLICT 382 382 Y -> F (IN REF. 3)
 CC -----

```

FT CONFLICT 501 501 I -> T (IN REF. 3).
FT CONFLICT 568 568 R -> G (IN REF. 3).
FT CONFLICT 570 570 T -> A (IN REF. 3).
FT CONFLICT 693 695 DIL -> VYT (IN REF. 4).
SQ SEQUENCE 1122 AA; 121305 MW; 155342DA2D6C3C65 CRC64;

Query Match
Best Local Similarity 4.5%; Score 110.5; DB 1; Length 1122;
Matches 94; Conservative 77; Mismatches 193; Indels 131; Gaps 26;

QY 42 TTPPGGTSFAGHNSAICASSAYINFGQPQNDYLVTPSLSPNGGTLTFWCAQDAN 101
DB 437 TAEAPGNTKVGYPYGLLSA---ISFDATR-----NGLALAPALQDVG 479
QY 102 YASEHYA-----VYASST--GNDASNFAN-----ALLSEVLTAKTV 135
DB 480 Y---HFVPLAVGGVSSPRGANGNIFLGSALTWTGNGGNFLDTKWHSPAVIEDAFTTIT 536
QY 136 VTAPEAIRGTRVQGTWYQKTVQLP--AGTKYVAFRHFGCTDFFWNLNDDVEIK---ANGK 190
DB 537 VNSSGVQNS---GQQSTSTPMPSNGNESIPYRTWNSYDYSVRFPAALISKPAGGNTK 593
QY 191 RAD--FTTFPSSTHGEAPAEWTTID-----ADGGQGWCLSGQLGWLTAHGQTNV 242
DB 594 QVESLFTALKDLTLNSLPNFTQENNIFFSVAMLDGQWLSLGTWKDSTWLT---TNTI 649
QY 243 ASFSWN-----GHALPNPNYLIIS---KDV-----TGATKVKYIYAV 275
DB 650 NNFTYNTQQQLASTAAGENANPRNLNALTAKGDFRDRDIGNVDILYSNNTKFTYYQV 709
QY 276 ND-----GPPGDHYAVMISKTGYNAGDFTVVEETENGINK--GGARFGSLSTEADG 324
DB 710 GGALTTPVEQVQNKTSANIYYMLTRDFGSTTPATODANTVSSKUNGAYLSSTGQQG 769
QY 325 AKPOSVMIERVLDLPAGTKYV--APRHYNCSDLNLYLLDDIQFTWGGSPPTDITYTYVR 382
DB 770 WYNGSIYVKKASFTSSQGYTWQDFKGLTTTASNAVISN---WTKAG-----YSIRP 818
QY 383 DGTIKKGLTETFEEDGVAICN-----HEYCVE-----VKYTAGVSPKCVNVTVPVQ 432
DB 819 DDDTV--FSVSKIPPEKEITAAVNRSLDSYVQLNGETSVNTVARVSPDSSA-LTLNPKR 876
QY 433 F-NPQVQLTGSAGVG 446
DB 877 ITNPLMN-RDNVICQ 890

RESULT 15
PME_ASFAC
ID PME_ASFAC STANDARD; PRF; 331 AA.
AC Q12535;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pectinesterase precursor (EC 3.1.1.11) (Pectin methylesterase) (PE).
GN PME1.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=KSM 510;
RX MEDLINE=97079238; PubMed=8920970;
RA Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA Doreich K., Dalboege H., Kauppinen S.;
RT "Pectin methyl esterase from Aspergillus aculeatus: expression
RT cloning in yeast and characterization of the recombinant enzyme.";
RL Biochem. J. 319:705-712 (1996).
CC -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC -!- SIMILARITY: Belongs to the pectinesterase family.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U49378; AAB42153.1; -
 DR InterPro; IPR000070; Pectinesterase.
 DR Pfam; PF01095; Pectinesterase; 1.
 DR PROSITE; PS00800; PECTINESTERASE_1; 1.
 DR PROSITE; PS00503; PECTINESTERASE_2; 1.
 KW Hydrolase; Aspartyl esterase; Cell wall; Signal.
 FT SIGNAL 1 17 BY SIMILARITY.
 FT CHAIN 18 331 PECTINESTERASE.
 FT ACT_SITE 162 162 BY SIMILARITY.
 FT ACT_SITE 183 183 BY SIMILARITY.
 SQ SEQUENCE 331 AA; 35681 MW; 1F1C81BF1E32174F CRC64;

Query Match
 Best Local Similarity 4.4%; Score 109.5; DB 1; Length 331;
 Matches 69; Conservative 33; Mismatches 99; Indels 105; Gaps 15;
 QY 108 AVYASSTGNDASNFANALLEEVLTAKTVTVAPEAIRGTRVQGTWYQKTVQLPAGT-KYVA 166
 DB 27 AIVVAKSGDVTITGDAL--DALSTSTDTQITIE-----EGT-YDEQVYLPAMTGKVII 79
 QY 167 FRHGGCTDFFWINL-----DDVEIKANGKRAFDTTFE-----SSTHGEAPAE 209
 DB 80 YGQTEHTDSYADNLVITTHAISYEDAGESDDLTAIFRNKAVGSOVYNLIANTCGQACHQ 139
 QY 210 WTTIDADGGQWGL-CLSSGQLGWLTAHGQTNV----- 242
 DB 140 ALALSADWQQGYGNCNFTGYQDTLLAQTNQLYINSYIEGAVDFIFGQHARAWFQVNDI 199
 QY 243 -----ASFWSNGMALNPDN--YLISKDVTGA-----TKVYIYAVNDGFGDHYAV 286
 DB 200 RVVEGPTASITANGRSSETDTSYVINKSTVAKEGDDVAEGTYL-----GRWSEYA- 254
 QY 287 MISKGTNAGDFTVVFEEET--PNCINKGARFGLSTEADGAKPOSVMIERVLDLPAGTKY 344
 DB 255 -----RVVFOQTSMNTVNSLIG-----WTEWSTSTP-NTEY 284
 QY 345 VAFRHY 350
 DB 285 VTFGEY 290

Search completed: May 18, 2004, 11:43:34
 Job time : 10.5342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:36:20 ; Search time 33.3368 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTPNPENPNPTTLSESF.....QNLTSAGVQKVLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | DB ID | Description |
|------------|--------|---------------|---------|-------------|
| 1 | 2094 | 84.4 | 1723 2 | P72197 |
| 2 | 2082 | 84.0 | 1687 2 | Q9R9B7 |
| 3 | 2082 | 84.0 | 1704 2 | Q51816 |
| 4 | 2079 | 83.8 | 1723 2 | P72194 |
| 5 | 2071 | 83.5 | 1358 2 | P96967 |
| 6 | 2070 | 83.5 | 1706 2 | Q51839 |
| 7 | 2059 | 83.0 | 1706 2 | Q51838 |
| 8 | 2054.5 | 82.8 | 1223 2 | Q9ZNB5 |
| 9 | 2039.5 | 82.2 | 1097 2 | P72196 |
| 10 | 2039.5 | 82.2 | 1732 2 | O07442 |
| 11 | 2034.5 | 82.0 | 1732 2 | Q51817 |
| 12 | 2020.5 | 81.5 | 1732 2 | O52050 |
| 13 | 682 | 27.5 | 925 2 | Q9F4J0 |
| 14 | 561.5 | 22.6 | 312 2 | Q9KIB3 |
| 15 | 233 | 9.4 | 293 2 | Q9XB09 |
| 16 | 152.5 | 6.1 | 1742 16 | Q9P377 |

| | | | | | | |
|----|-------|-----|------|----|--------|---------------------|
| 17 | 144 | 5.8 | 3346 | 16 | Q7WN54 | Q7wn54 bordetella |
| 18 | 139.5 | 5.6 | 1341 | 16 | Q8UAU1 | Q8uau1 agrobacteri |
| 19 | 139 | 5.6 | 2215 | 16 | Q7WBN0 | Q7wb0 bordetella |
| 20 | 135 | 5.4 | 3988 | 17 | Q8TFZ1 | Q8tfx1 methanosarc |
| 21 | 134.5 | 5.4 | 2768 | 16 | Q8E9G6 | Q8e9g6 shewanella |
| 22 | 134 | 5.4 | 1541 | 16 | Q7UQJ9 | Q7uqj9 rhodopirell |
| 23 | 134 | 5.4 | 1800 | 2 | Q9L948 | Q9l948 pseudomonas |
| 24 | 134 | 5.4 | 8682 | 16 | Q88RG2 | Q88rg2 pseudomonas |
| 25 | 133.5 | 5.4 | 680 | 2 | O52644 | O52644 ruminococcu |
| 26 | 133 | 5.4 | 2522 | 16 | Q8EKA6 | Q8ek6 shewanella |
| 27 | 132.5 | 5.3 | 555 | 16 | Q82EW1 | Q82ew1 streptomyce |
| 28 | 132.5 | 5.3 | 691 | 16 | Q9RZS7 | Q9rzs7 deinococcus |
| 29 | 132.5 | 5.3 | 1115 | 4 | Q9HD43 | Q9hd43 homo sapien |
| 30 | 131.5 | 5.3 | 1256 | 16 | Q8NX96 | Q8nx96 staphylococ |
| 31 | 131.5 | 5.3 | 1357 | 17 | Q8THC8 | Q8thc8 methanosarc |
| 32 | 131.5 | 5.3 | 2468 | 16 | Q912W3 | Q912w3 pseudomonas |
| 33 | 131 | 5.3 | 2219 | 16 | Q88W19 | Q88w19 lactobacill |
| 34 | 130.5 | 5.3 | 880 | 2 | Q9RMB8 | Q9rmb8 arthrobacte |
| 35 | 130.5 | 5.3 | 955 | 17 | Q8TQ91 | Q8tg91 methanosarc |
| 36 | 130 | 5.2 | 1386 | 17 | Q8TI72 | Q8ti72 methanosarc |
| 37 | 129.5 | 5.2 | 756 | 9 | Q858B6 | Q858b6 enterobacte |
| 38 | 129.5 | 5.2 | 1744 | 16 | Q8JY88 | Q8jy88 enterococcu |
| 39 | 129 | 5.2 | 7716 | 16 | Q7UMZ8 | Q7umz8 rhodopirell |
| 40 | 128.5 | 5.2 | 1248 | 16 | Q99V41 | Q99v41 staphylococ |
| 41 | 128.5 | 5.2 | 1255 | 2 | Q7WTC6 | Q7wtc6 staphylococ |
| 42 | 128.5 | 5.2 | 2656 | 5 | Q9GNU3 | Q9gnu3 paracentrot |
| 43 | 128 | 5.2 | 1672 | 16 | Q8Y366 | Q8y366 ralatonia s |
| 44 | 128 | 5.2 | 4936 | 16 | Q8YKJ3 | Q8ykj3 anabaena sp |
| 45 | 127.5 | 5.1 | 1684 | 2 | Q03658 | Q03658 unidentified |

ALIGNMENTS

RESULT 1

| | | | | | |
|--------|---|---|--------------|------|----------|
| P72197 | AC | P72197 | PRELIMINARY; | PRT; | 1723 AA. |
| DT | 01-FEB-1997 | (T-EMBLrel. 02, Created) | | | |
| DT | 01-FEB-1997 | (T-EMBLrel. 02, Last sequence update) | | | |
| DT | 01-OCT-2003 | (T-EMBLrel. 25, Last annotation update) | | | |
| DE | lys-gingipain. | | | | |
| GN | KGP. | | | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales; | | | | |
| OC | Porphyromonadaceae; Porphyromonas. | | | | |
| OX | NCBI_TaxID=837; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N., | | | | |
| RA | Prochazka V., Kiefer M.C., Travis J., Barr P.J.; | | | | |
| RT | "Molecular cloning and characterization of Porphyromonas gingivalis | | | | |
| RT | Lys-gingipain." | | | | |
| RL | Submitted (APR-1996) to the EMBL/GenBank/DBJ databases. | | | | |
| DR | EMBL; U54691; AAA99810.1; -. | | | | |
| DR | MEROPS; C25.002; -. | | | | |
| DR | GO; GO:0005524; F:ATP binding; IEA. | | | | |
| DR | GO; GO:0008234; F:cysteine-type peptidase activity; IEA. | | | | |
| DR | GO; GO:0003910; F:DNA ligase (ATP) activity; IEA. | | | | |
| DR | GO; GO:0006310; F:DNA recombination; IEA. | | | | |
| DR | GO; GO:0006281; F:DNA repair; IEA. | | | | |
| DR | GO; GO:0006260; F:DNA replication; IEA. | | | | |
| DR | GO; GO:0006508; P:proteolysis and peptidolysis; IEA. | | | | |
| DR | InterPro; IPR001977; DNA ligase. | | | | |
| DR | InterPro; IPR001769; Peptidase_C25. | | | | |
| DR | InterPro; IPR005536; Peptidase_C25. | | | | |
| DR | Pfam; PF01364; Peptidase_C25; 1. | | | | |
| DR | Pfam; PF03785; Peptidase_C25; 1. | | | | |
| DR | PROSITE; PS00697; DNA_LIGASE_A1; 1. | | | | |
| SQ | SEQUENCE 1723 AA; 166831 MW; 4508A7B50197CEED CRC64; | | | | |

Query Match

84.4%

Score 2094;

DB 2;

Length 1723;

Best Local Similarity

86.2%

Pred. No. 6.8e-128;

Matches 394; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 1 GTNPENPNPGTITLSEFENGIPASWKITIDADGCGNNWTTTPPGGTSPAGHNSAICA 60
 Db 969 GTNPENPNPGTITLSEFENGIPASWKITIDADGCGHGWKPGNAPG---IAGYNSGCV 1025

QY 61 SSASY-INFGPQPNPNVLTPELSLNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
 Db 1026 YSEFGLGGIGVLTPTDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

QY 120 NEANALLEEVLTAKTVTAPAIRGTRVQGTWYOKTVOLPAGTKYVAFRHFQCTDFFWIN 179
 Db 1086 NFNALLEETITAG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYID 1143

QY 180 LDVEIKANGKRAADFTETTFESSHGEAPAEWTTIDADGCGQWMLCLSSGQLWLTAGGT 239
 Db 1144 LDVEIKANGKRAADFTETTFESSHGEAPAEWTTIDADGCGQWMLCLSSGQLWLTAGGT 1203

QY 240 NVVASTSWNGMALPNPNYLISKDVGTGATKVKYIYAVNDGPGDHYAVMISKTGTNAGDFT 299
 Db 1204 NVVASTSWNGMALPNPNYLISKDVGTGATKVKYIYAVNDGPGDHYAVMISKTGTNAGDFT 1263

QY 300 VFEETPNGLKNGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 359
 Db 1264 VFEETPNGLKNGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 1323

QY 360 LDDIQTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGATGNHCEYCVKVTAGVS 419
 Db 1324 LDDIQTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGATGNHCEYCVKVTAGVS 1383

QY 420 PKECNVNVTDPQPNVQNLTSAGVQKVLKWDAPN 456
 Db 1384 PKECNVNVTDPQPNVQNLTSAGVQKVLKWDAPN 1420

RESULT 2
 Q9R9B7 PRELIMINARY; PRT; 1687 AA.

AC Q9R9B7
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hemagglutinin/protease.
 GN HAGE
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RA Han N., Dong H., Progleke-Fox A.;
 RT "Cloning and characterization of hage from P. gingivalis 381.";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF026946; AA01810.1; -;
 DR HSP; P95493; 1CVR.
 DR MEROPS; C25.001; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase C25.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64;

Query Match 84.0%; Score 2082; DB 2; Length 1687;
 Best Local Similarity 86.2%; Pred. No. 4e-127;
 Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPNPNPNPGTITLSEFENGIPASWKITIDADGCGNNWTTTPPGGTSPAGHNSAICA 62
 Db 935 PNPNPNPNPGTITLSEFENGIPASWKITIDADGCGHGWKPGNAPG---IAGYNSGCVTS 991

QY 63 ASY-INFGPQPNPNVLTPELSLNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
 Db 992 ESFGLGGIGVLTPTDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

QY 122 ANALLEEVLTAKTVTAPAIRGTRVQGTWYOKTVOLPAGTKYVAFRHFQCTDFFWINLD 181
 Db 1052 NFNALLEETITAG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLD 1109

QY 182 LDVEIKANGKRAADFTETTFESSHGEAPAEWTTIDADGCGQWMLCLSSGQLWLTAGGTNV 241
 Db 1110 LDVEIKANGKRAADFTETTFESSHGEAPAEWTTIDADGCGQWMLCLSSGQLWLTAGGTNV 1169

QY 242 VASFSWNGMALPNPNYLISKDVGTGATKVKYIYAVNDGPGDHYAVMISKTGTNAGDFTV 301
 Db 1170 VASFSWNGMALPNPNYLISKDVGTGATKVKYIYAVNDGPGDHYAVMISKTGTNAGDFTV 1229

QY 302 FEETPNGLKNGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 361
 Db 1230 FEETPNGLKNGGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 1289

QY 362 DIQFTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGATGNHCEYCVKVTAGVSPK 421
 Db 1290 DIQFTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGATGNHCEYCVKVTAGVSPK 1349

QY 422 ECNVNVTDPQPNVQNLTSAGVQKVLKWDAPN 456
 Db 1350 ECNVNVTDPQPNVQNLTSAGVQKVLKWDAPN 1384

RESULT 3
 Q51816 PRELIMINARY; PRT; 1704 AA.

AC Q51816
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Arg-gingipain-1, proteinase.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95138080; PubMed=7836351;
 RA Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
 RA Travis J., Barr P.J.;
 RT "Molecular cloning and structural characterization of the Arg-
 RT gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a
 RL J. Biol. Chem. 270:1007-1010(1995).
 DR EMBL; U15282; AAA69539.1; -;
 DR FIR; A55426; A55426.
 DR HSP; P95493; 1CVR.
 DR MEROPS; C25.001; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase C25.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64;

DR GO: 0006310; P: DNA recombination; IEA.
 DR GO: 0006281; P: DNA repair; IEA.
 DR GO: 0006260; P: DNA replication; IEA.
 DR GO: 0006508; P: proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000977; DNA ligase.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001769; Peptidase C25.
 DR InterPro: IPR005536; Peptidase C25_C.
 DR Pfam: PF01364; Peptidase C25; I.
 DR Pfam: PF03785; Peptidase C25_C; 1.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

Query Match 83.5%; Score 2071; DB 2; Length 1358;
 Best Local Similarity 85.6%; Pred. No. 1.6e-126;
 Matches 391; Conservative 17; Mismatches 43; Indels 6; Gaps 4;

QY 1 GTNPENPNPGTTTISESPENGIPASWKITDADGQNNWTTTTPPGGTGFAGHNSAICA 60
 DB 604 GTNPENPNPGTTTISESPENGIPASWKITDADGQNNWTTTTPPGGTGFAGHNSAICA 60

QY 61 SSASY-INPEGPQNPNDYLTPELSLPGGTLFWVCAQDANYASEHYAVYASSTGNDAS 119
 DB 661 YSEFGLGGIGVLTDPDNYLTLPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDAS 119

QY 120 NFANALLEVLTAKTVVTAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQGTDFFWIN 179
 DB 721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWQKTVQLPAGTKYVAFRHFQGTDFFWIN 179

QY 180 LDDVEIKANGKRADEFTEFESSHGEAPAEWTTIDADGQNNWTTTTPPGGTGFAGHNSAICA 239
 DB 779 LDEVEIKANGKRADEFTEFESSHGEAPAEWTTIDADGQNNWTTTTPPGGTGFAGHNSAICA 239

QY 240 NVVASFSGWGMALPNPNYLISKDVTGATKYKYVAVNDGPPGDHYAVMISKTGTNAGDFT 299
 DB 839 NVVASFSGWGMALPNPNYLISKDVTGATKYKYVAVNDGPPGDHYAVMISKTGTNAGDFT 299

QY 300 VFEETPENGKNGARFGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHFQGTDFFWIN 359
 DB 899 VFEETPENGKNGARFGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHFQGTDFFWIN 359

QY 360 LDDIQTMGSPPTDITYVVRDGTGKIKEGLTETFEEDGVATGNHCEVVEKVTAGVS 419
 DB 959 LDDIQTMGSPPTDITYVVRDGTGKIKEGLTETFEEDGVATGNHCEVVEKVTAGVS 419

QY 420 PKECVNVTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 456
 DB 1019 PKVCVNVTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1055

RESULT 6

Q51839 PRELIMINARY; PRT; 1706 AA.
 AC Q51839; Q51840;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arginine-specific thiol protease precursor.
 GN PRTR.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RX MEDLINE=95160709; PubMed=7857299;
 RA Kirsbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
 RA Reynolds E.C.;
 RT "Complete nucleotide sequence of a gene prtr of Porphyromonas
 RT gingivalis W50 encoding a 132 kDa protein that contains an arginine-
 RT specific thiol endopeptidase domain and a haemagglutinin domain.";

RL Biochem. Biophys. Res. Commun. 207:424-431(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RX MEDLINE=96311339; PubMed=8713096;
 RA Slakeski N., Cleal S.M., Reynolds E.C.;
 RT "Characterization of a Porphyromonas gingivalis gene prtr that encodes
 RT an arginine-specific thiol proteinase and multiple adhesins.";
 RL Biochem. Biophys. Res. Commun. 224:605-610(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Reynolds E.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Slakeski N.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L26341; AAC18876.1; -.
 DR HSP; P35493; ICVR.
 DR GO: 0005524; F: ATP binding; IEA.
 DR GO: 0008234; F: cysteine-type peptidase activity; IEA.
 DR GO: 0003910; F: DNA ligase (ATP) activity; IEA.
 DR GO: 0006310; P: DNA recombination; IEA.
 DR GO: 0006281; P: DNA repair; IEA.
 DR GO: 0006260; P: DNA replication; IEA.
 DR GO: 0008508; P: proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000977; DNA ligase.
 DR InterPro: IPR007110; IG-Like.
 DR InterPro: IPR001769; Peptidase C25.
 DR InterPro: IPR005536; Peptidase C25_C.
 DR Pfam: PF01364; Peptidase C25; I.
 DR Pfam: PF03785; Peptidase C25_C; 1.
 DR PROSITE: PS00697; DNA_LIGASE_A1; 1.
 KW Protease; Signal.
 FT SIGNAL 1 23 POTENTIAL,
 FT CHAIN 228 1706 ARGinine-SPECIFIC THIOIOL PROTEASE.
 SQ SEQUENCE 1706 AA; 185626 MW; 88BDF07C9813B844 CRC64;

Query Match 83.5%; Score 2070; DB 2; Length 1706;
 Best Local Similarity 85.7%; Pred. No. 2.5e-126;
 Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTISESPENGIPASWKITDADGQNNWTTTTPPGGTGFAGHNSAICA 62
 DB 954 PNPENPNPGTTTISESPENGIPASWKITDADGQNNWTTTTPPGGTGFAGHNSAICA 62

QY 63 ASY-INPEGPQNPNDYLTPELSLPGGTLFWVCAQDANYASEHYAVYASSTGNDASNF 121
 DB 1011 ESFGLGGIGVLTDPDNYLTLPALDLPNGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1010

QY 122 ANALLEVLTAKTVVTAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQGTDFFWINLD 181
 DB 1071 TNALLEETITAKG-VRSPEAIRG-RIQGTWQKTVQLPAGTKYVAFRHFQGTDFFWINLD 1128

QY 182 DVEIKANGKRADEFTEFESSHGEAPAEWTTIDADGQNNWTTTTPPGGTGFAGHNSAICA 241
 DB 1129 EVEIKANGKRADEFTEFESSHGEAPAEWTTIDADGQNNWTTTTPPGGTGFAGHNSAICA 1188

QY 242 VASFSGWGMALPNPNYLISKDVTGATKYKYVAVNDGPPGDHYAVMISKTGTNAGDFTV 301
 DB 1189 VASFSGWGMALPNPNYLISKDVTGATKYKYVAVNDGPPGDHYAVMISKTGTNAGDFTV 1248

QY 302 FEETPENGKNGARFGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHFQGTDFFWINLD 361
 DB 1249 FEETPENGKNGARFGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHFQGTDFFWINLD 1308

QY 362 DIQTMGSPPTDITYVVRDGTGKIKEGLTETFEEDGVATGNHCEVVEKVTAGVSPK 421
 DB 1309 DIQTMGSPPTDITYVVRDGTGKIKEGLTETFEEDGVATGNHCEVVEKVTAGVSPK 1368

QY 422 ECNVNVTVDVPQFNPVQNLTSAGVQKVTWKWDAPN 456

| | | | |
|----------|--|--|-----------------|
| Db | 1369 | KCNVTVNSTQFNPNVKNLKAQPDGSDVVLKWEAPS | 1403 |
| | | : : : : : : : | |
| RESULT 7 | | | |
| Q51838 | | | |
| ID | Q51838 | PRELIMINARY; | PRT; 1706 AA. |
| AC | Q51838; | | |
| DT | 01-NOV-1996 | (TrEMBLrel. 01, Created) | |
| DT | 01-MAY-1997 | (TrEMBLrel. 03, Last sequence update) | |
| DT | 01-OCT-2003 | (TrEMBLrel. 25, Last annotation update) | |
| DE | Protease precursor. | | |
| GN | PAPRI. | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales; | | |
| OC | Porphyromonadaceae; Porphyromonas. | | |
| OX | NCBI_TaxID=837; | | |
| RN | [1] | SEQUENCE FROM N.A. | |
| RC | STRAIN=W50; | | |
| RC | STRAIN=W50; | | |
| RX | MEDLINE=96071894; PubMed=7591131; | | |
| RA | Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.; | | |
| RT | "Characterization, Genetic analysis, and expression of a protease | | |
| RT | antigen (PpRI) of Porphyromonas gingivalis W50."; | | |
| RL | Infect. Immun. 63:4744-4754(1995). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=W50; | | |
| RA | Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.; | | |
| RT | "The pprI and the ppr2 arginine-specific protease genes of | | |
| RT | Porphyromonas gingivalis W50 produce five biochemically distinct | | |
| RT | enzymes."; | | |
| RL | Mol. Microbiol. 23:0-0(1997). | | |
| DR | EMBL; X82680; CAA57997.1; -. | | |
| DR | HSSP; P95493; 1CVR. | | |
| DR | MEROPO; C25.001; -. | | |
| DR | GO; GO:0005524; F:ATP binding; IEA. | | |
| DR | GO; GO:0008234; F:cysteine-type peptidase activity; IEA. | | |
| DR | GO; GO:0003910; F:DNA ligase (ATP) activity; IEA. | | |
| DR | GO; GO:0006310; F:DNA recombination; IEA. | | |
| DR | GO; GO:0006281; F:DNA repair; IEA. | | |
| DR | GO; GO:0006260; F:DNA replication; IEA. | | |
| DR | GO; GO:0006508; F:proteolysis and peptidolysis; IEA. | | |
| DR | InterPro; IPR000977; DNA ligase. | | |
| DR | InterPro; IPR007110; Ig-like. | | |
| DR | InterPro; IPR001769; Peptidase C25. | | |
| DR | InterPro; IPR005536; Peptidase C25. | | |
| DR | Pfam; PF01364; Peptidase C25; 1. | | |
| DR | Pfam; PF03785; Peptidase C25; 1. | | |
| DR | PROSITE; PS00697; DNA_LIGASE_A1; 1. | | |
| DR | Signal. | | |
| FT | SIGNAL | 1 23 | POTENTIAL. |
| FT | CHAIN | 228 719 | ALPHA-PROTEASE. |
| FT | CHAIN | 720 1262 | BETA-ADHESIN. |
| SQ | SEQUENCE | 1706 AA; 185705 MW; OE56DCD87FDA8CDD | CRC64; |
| | Query Match | 83.0%; Score 2059; DB 2; Length 1706; | |
| | Best Local Similarity | 85.5%; Pred. No. 1.3e-125; | |
| | Matches 389; Conservative | 17; Mismatches 43; Indels 6; Gaps 4; | |
| QY | 3 | PNNFNPNGTTLTSESFENGIPASWKTIADGDNKWTTPPGGTGFACHNSAICASS | 62 |
| DB | 954 | PNNFNPNGTTLTSESFENGIPASWKTIADGDNKWTTPPGGTGFACHNSAICASS | 1010 |
| QY | 63 | ASV-INTEGPNPNLYVTPELSLPNGGTLTFWVCAQDANYASHYAVYASSTGNDASNF | 121 |
| DB | 1011 | ESFCLGIGVLTDPNLYLTALDPLNGSKLTFWVCAQDANYASHYAVYASSTGNDASNF | 1070 |
| QY | 122 | ANALLEVLTAQTVTAPEAIRGTRVGQWYKTVLPAGTKYVAFRHFGCTGFFWINLD | 181 |
| DB | 1071 | TNALLEETITAKG-VRSPEAIRG-RIQSTWRQKTVLDPAGTKYVAFRHFGCTGFFWINLD | 1128 |
| QY | 182 | DVEIKANGKRAQDFTETPESSTHGEAPAWTTIADGGGQGWLCSSQGLWLTFAHGCTNV | 241 |


```

QY 179 NLDDVEIKANGKADTFETFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 238
Db 635 DLDEVEIKANGKADTFETFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 694
QY 239 TNVVSFSWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 298
Db 695 TNVVSFSWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 754
QY 299 TVVFEETPGNGKKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHYNCSDLNYI 358
Db 755 TVVFEETPGNGKKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHYNCSDLNYI 814
QY 359 LLDDIQFTMGSSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGV 418
Db 815 LLDDIQFTMGSSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGV 874
QY 419 SPKVCNVTVDPVQFNPNVQNLTGSAV--GQKVLTKWDAP 455
Db 875 SPKVCNVTVDPVQFNPNVQNLTGSAV--GQKVLTKWDAP 913

RESULT 9
P72196
ID P72196 PRELIMINARY; PRT; 1097 AA.
AC P72196;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TonB-linked adhesin precursor.
GN TLA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97386416; PubMed=9244265;
RA Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
RA Curtis M.A.;
RT "The cla gene of Porphyromonas gingivalis W50: a homologue of the
RT arginine-specific protease precursor (PrpRI) which shares sequence
RT similarity to TonB-linked receptors."
RL J. Bacteriol. 179:4778-4788(1997).
CC 1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Y07618; CAA68897.1;
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000977; DNA ligase.
DR Pfam; PF00593; TonB dep Rec; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 53 POTENTIAL.
SQ SEQUENCE 1097 AA; 118731 MW; 73BBA337B421F8B9 CRC64;

Query Match 82.2%; Score 2039.5; DB 2; Length 1097;
Best Local Similarity 84.5%; Pred. No. 1.3e-124;
Matches 388; Conservative 18; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPGT-TTLSEFENGIPASWKTIDADGNNWTTTPPGCTGFAGHNSAIC 59
Db 334 GTPNPNPNPNPGTTLSEFENGIPASWKTIDADGNNWTTTPPGCTGFAGHNSAIC 390
QY 60 ASSAY- INFEGQPNPNVLPTELSPNGGTLTFWVCAQDANYASEHYAVYSTGND 118

```

```

Db 391 VYSEFGLGGIGVLTDPNLYITPALDLPNGKLTFFWCAQDANYASEHYAVYSTGND 450
QY 119 SNFANALLEVLTAKTAVVTAPEAIRTRVGQVTKTVQVLPAGTKYVAFRHFQCTDFWI 178
Db 451 SNTFNALLEETITAKG-VRSFKAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFYI 508
QY 179 NLDDVEIKANGKADTFETFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 238
Db 509 DLDEVEIKANGKADTFETFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 568
QY 239 TNVVSFSWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 298
Db 569 SNTVVSFSWNGMALPNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 628
QY 299 TVVFEETPGNGKKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHYNCSDLNYI 358
Db 629 TVVFEETPGNGKKGARFGLSTEADGAKPQSVWIERVTDLPAGTKYVAFRHYNCSDLNYI 688
QY 359 LLDDIQFTMGSSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGV 418
Db 689 LLDDIQFTMGSSPTPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKYTAGV 748
QY 419 SPKVCNVTVDPVQFNPNVQNLTGSAV--GQKVLTKWDAP 455
Db 749 SPKVCNVTVDPVQFNPNVQNLTGSAV--GQKVLTKWDAP 787

RESULT 10
O07442
ID O07442 PRELIMINARY; PRT; 1732 AA.
AC O07442;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lysine-specific cysteine proteinase.
GN PRK.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=W50;
RX MEDLINE=99235907; PubMed=10219167;
RA Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
RT "Characterization of a Porphyromonas gingivalis gene prk that encodes
RT a lysine-specific cysteine proteinase and three sequence-related
RT adhesins."
RL Oral Microbiol. Immunol. 14:92-97(1999).
DR EMBL; U75366; AAB60809.1;
DR MEROPS; C25.002;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1732 AA; 187914 MW; 45D5B91377391703 CRC64;

Query Match 82.2%; Score 2039.5; DB 2; Length 1732;
Best Local Similarity 84.5%; Pred. No. 2.4e-124;
Matches 388; Conservative 18; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPGT-TTLSEFENGIPASWKTIDADGNNWTTTPPGCTGFAGHNSAIC 59
Db 969 GTPNPNPNPNPGTTLSEFENGIPASWKTIDADGNNWTTTPPGCTGFAGHNSAIC 1025

```

QY 60 ASSASY-INTEGPNPNPNLTPELSPNGGTLTFWVCAQDANYASEHYAVASSTGND 118
 Db 1026 VYSESFLGGIGVLTDPNLTLPALDLPNGGKLTFFWCAQDANYASEHYAVASSTGND 1085
 QY 119 SNFANALLEEVLTAKTAVTAPALRGTRVQGTWQKTVQVLPAGTKYVAFRHFQCTDFWI 178
 Db 1086 SNFTNALLEETITAKG-VRSFKAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQCTDMFYI 1143
 QY 179 NLDDVEIKANGKRAADTFETFESSTHGAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 238
 Db 1144 DLDEVEIKANGKRAADTFETFESSTHGAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 1203
 QY 239 TNNVASFWSNGMALPNPNYLISKDVTGATKYYAVVNDGPGDHYAVMISKTGTNAGDF 298
 Db 1204 SNVVSFSWSNGMALPNPNYLISKDVTGATKYYAVVNDGPGDHYAVMISKTGTNAGDF 1263
 QY 299 TVVFEETPNNGKNGARFGLSTEADGAKPOSVWERTVLDLPAGTKYVAFRHYNCSDLNYI 358
 Db 1264 TVVFEETPNNGKNGARFGLSTEADGAKPOSVWERTVLDLPAGTKYVAFRHYNCSDLNYI 1323
 QY 359 LLDDIQFTMGSSPTPTDYTVVYRDGTKIKEGLTETTFEEDGVATGNHHEYCVVKYTAGV 418
 Db 1324 LLDDIQFTMGSSPTPTDYTVVYRDGTKIKEGLTETTFEEDGVATGNHHEYCVVKYTAGV 1383
 QY 419 SPKECVNVTVPQFNPVQNLTGSAV--GOKVTLKWDAP 455
 Db 1384 SPKCKVDVTNVTQFNPVQNLTAQAPNSMDAILKWNAP 1422

RESULT 11

Q51817 PRELIMINARY; PRT; 1732 AA.
 AC Q51817
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Porphyain.
 GN PRTP.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=W12;
 RX MEDLINE=96213011; PubMed=8631659;
 RA Barkocy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
 RA Progulskie-Fox A., Lantz M.S.;
 RT "Analysis of the prtp gene encoding porphyain, a cysteine proteinase
 of Porphyromonas gingivalis";
 RL J. Bacteriol. 178:2734-2741(1996).
 DR EMBL; U42210; AAB06565.1; -.
 DR F01; T30836; T30836.
 DR MEROPS; C25.002; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase C25_C.
 DR Pfam; PF01364; Peptidase C25; 1.
 DR Pfam; PF03785; Peptidase C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCAB4 CRC64;

Query Match

Best Local Similarity 82.0%; Score 2034.5; DB 2; Length 1732;
 Matches 387; Conservative 19; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPGT--TTLSSEFENGIPASWKTIDADGGNNWTTTPPGCTSPFAGNSAIC 59
 Db 969 GTPNPNPNPNPGT--TTLSSEFENGIPASWKTIDADGGNNWTTTPPGCTSPFAGNSAIC 1025
 QY 60 ASSASY-INTEGPNPNPNLTPELSPNGGTLTFWVCAQDANYASEHYAVASSTGND 118
 Db 1026 VYSESFLGGIGVLTDPNLTLPALDLPNGGKLTFFWCAQDANYASEHYAVASSTGND 1085
 QY 119 SNFANALLEEVLTAKTAVTAPALRGTRVQGTWQKTVQVLPAGTKYVAFRHFQCTDFWI 178
 Db 1086 SNFTNALLEETITAKG-VRSFKAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQCTDMFYI 1143
 QY 179 NLDDVEIKANGKRAADTFETFESSTHGAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 238
 Db 1144 DLDEVEIKANGKRAADTFETFESSTHGAPAEWTTIDADGGQGWCLSSGQGLWLTAGG 1203
 QY 239 TNNVASFWSNGMALPNPNYLISKDVTGATKYYAVVNDGPGDHYAVMISKTGTNAGDF 298
 Db 1204 SNVVSFSWSNGMALPNPNYLISKDVTGATKYYAVVNDGPGDHYAVMISKTGTNAGDF 1263
 QY 299 TVVFEETPNNGKNGARFGLSTEADGAKPOSVWERTVLDLPAGTKYVAFRHYNCSDLNYI 358
 Db 1264 TVVFEETPNNGKNGARFGLSTEADGAKPOSVWERTVLDLPAGTKYVAFRHYNCSDLNYI 1323
 QY 359 LLDDIQFTMGSSPTPTDYTVVYRDGTKIKEGLTETTFEEDGVATGNHHEYCVVKYTAGV 418
 Db 1324 LLDDIQFTMGSSPTPTDYTVVYRDGTKIKEGLTETTFEEDGVATGNHHEYCVVKYTAGV 1383
 QY 419 SPKECVNVTVPQFNPVQNLTGSAV--GOKVTLKWDAP 455
 Db 1384 SPKCKVDVTNVTQFNPVQNLTAQAPNSMDAILKWNAP 1422

RESULT 12

Q52050 PRELIMINARY; PRT; 1732 AA.
 AC Q52050
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Lysine specific cysteine protease.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=W83;
 RX MEDLINE=98298016; PubMed=9632563;
 RA Lewis J.P., Macrina F.L.;
 RT "IS195, an insertion sequence-like element associated with protease
 genes in Porphyromonas gingivalis";
 RL Infect. Immun. 66:3035-3042(1998).
 DR EMBL; AF017059; AAC26523.1; -.
 DR MEROPS; C25.002; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase C25_C.
 DR Pfam; PF01364; Peptidase C25; 1.
 DR Pfam; PF03785; Peptidase C25; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1732 AA; 187931 MW; B2337463D5C85EA5 CRC64;
 Query Match 81.5%; Score 2020.5; DB 2; Length 1732;

```

Best Local Similarity 83.9%; Pred. No. 4.2e-123;
Matches 385; Conservative 19; Mismatches 46; Indels 9; Gaps 6;

QY 1 GTPNPNPNPGT-TTISESPENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAIC 59
Db 969 GTPNPNPNPNPGT-TTISESPENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAIC 1025
QY 60 ASSASY-INTEGPNPNPNPGT-TTISESPENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAIC 118
Db 1026 YSSEFLGGIGVLTDPNYLITPALDLPNGGKLTFFWCAQDANYASEHYASSTGND 1085
QY 119 SNFANALLEEVLTAQVTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFQCTDFFWI 178
Db 1086 SNFTNALLEEITIAKG-VRPKAIKG-RIQGTWQKTVQVLPAGTKYVAFRHFQCTDFFWI 1143
QY 179 NLDDVEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGQCGWCLSSQGLWLTAGG 238
Db 1144 DLDEVEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGQCGWCLSSQGLWLTAGG 1203
QY 239 TTVVASFSGMNLNPNYLISKDVTGATKVKYKYYAVNDGPGDHYAVMISKTGTNAGDF 298
Db 1204 SNVVSFSWGMNLNPNYLISKDVTGATKVKYKYYAVNDGPGDHYAVMISKTGTNAGDF 1263
QY 299 TTVVFEETPNKGGARFGLSTADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDNYI 358
Db 1264 TTVVFEETPNKGGARFGLSTADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDNYI 1323
QY 359 LLDDIQTMGGSPTDPTDVTYVVRDGTNIKEGLTETTFEYVATGNHCEYVETAGV 418
Db 1324 LLDDIQTMGGSPTDPTDVTYVVRDGTNIKEGLTETTFEYVATGNHCEYVETAGV 1383
QY 419 SPKECVNVTVPQENPVQNLGSAV--GQVTLKWDAP 455
Db 1384 SPKCVNVTVPQENPVQNLGSAV--GQVTLKWDAP 1422

RESULT 13
Q9F4J0 PRELIMINARY; PRT; 925 AA.
AC Q9F4J0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Putative outer membrane protein PG57.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT "Identification of vaccine candidates from genomic analysis of
RT Porphyromonas gingivalis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007522; BAG24228.1;
DR InterPro; IPR003961; FN.III.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 925 AA; 103632 MW; 5FF2198D6914DAE8 CRC64;

Query Match 27.5%; Score 682; DB 2; Length 925;
Best Local Similarity 30.2%; Pred. No. 4e-36;
Matches 176; Conservative 62; Mismatches 132; Indels 212; Gaps 22;

QY 1.6 LSESPENG-IPASWKTIDADGDNWNTTTPPGG-----TSFAGHNSAICASSASYINF 68

```

```

Db 311 LYENFENGVPNGWLVADADGDFSW-----GHYLNAYDAFEGHGHCHLSASIVPG 363
QY 69 EGQNPNDNYLVTPELSIPNGGTLTFWVCAQDANYASEHYAVASSTGNDASNFANALLEE 128
Db 364 IGPTPDNYLITPKVE--GAKRVKYVWSTQDANAAEAHYAVMASTTGTAVGDFV-ILFEE 420
QY 129 VLTAKTVTAPEAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFQCTDFFWINLDDV---- 183
Db 421 TMTAKT-----GAWVERTINLPEGTKIAWRHYNCTDIYFLKDDITVFGT 467
QY 184 -----EIKANG----- 189
Db 468 PASEPEPVTDFVVSLLIENKGRKLNYPNGYEPDKTDKDLQLAGVNIYANGSLLVHIQ 527
QY 190 -----KRADEFTET-----FESSTHGE 205
Db 528 DFTVLEYIDETYSRDDQVEVEYCVTAVYNDNIESQVCDKLIYDSQSDIILYEGFEAGS 587
QY 206 APAEWTTIDADGQCGWCLSSQGLGW-LTAHGTTNVASFWSWGM--ALNPDNYLISKD 262
Db 588 IPEGWLLIDADGDNVNM-----DYYPWTWYGHDSKCIASPSYLPIMGVLPDNYLVTFR 642
QY 263 VTGATKVKYKYYAVNDG--FPGDHYAVMISKTGTNAGDFTVVFETPNKGGARFGLSTE 321
Db 643 LEGAKLVKYVSAQDAVYGAHYAVMVSTGTAVDFVLLFEET-----MTAK 690
QY 322 ADGAKPQSVIERTVDLEAGTKYVAFRHYNCSDNLVILLDDI----- 363
Db 691 ANGA-----WYERTITLPAGTKYIAWRHYDCTDMPFLLDDITVYRSTETPEPVTDFV 745
QY 364 -----QFTMGGSPTPTD-----VYTVYVRDGT-----KIKEGLTETTF 396
Db 746 SLIENKGRKLNYPNGYEPDKTDKDLQLAGVNIYANGSLLVHIQDFTVLEYIDETYS 805
QY 397 EEDGATGNHCEYVETAGV-SPKEC--VNVTV---DPVQ 432
Db 806 SRDGV--EMEYCVTAVYNDNIESQVCDKLNVTITSLDNIQ 845

RESULT 14
Q9KIB3 PRELIMINARY; PRT; 312 AA.
AC Q9KIB3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237555; AAF81413.1;
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match 22.6%; Score 561.5; DB 2; Length 312;
Best Local Similarity 32.8%; Pred. No. 7e-29;
Matches 136; Conservative 23; Mismatches 73; Indels 183; Gaps 8;

QY 1.6 LSESPENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAICASSASYINFQPNPD 75

```

```
Db      4 LSESESGIPAIWKIIDADGDGYNMHL-----TNFTG-QSGLCVSSASYIGVGALTPD 57
QY      76 NYLVTPELSPNGG--TLTFWYCAQDANYASHYAYASSTGNDASNFANALLEEVLAK 133
Db      58 NYLITPELKLPTDALVEIYVWCTQDLTAPSEHYAVYSSSTGNNAADFNLLYEETLAK 117
QY      134 TVVTAPAIRTRVGTWYQKTVQIPAGTKYVAFRHFGCTDPFNLDDVEIKANGKRAAD 193
Db      118 R-IOSPELIRGNRTQGVYQKVLNPNDTKYVAFRHFNSDNFNLNDEVSL----- 169
QY      194 FTETPESSTHGEAPAEWTTIDADGGQGWCLSSQLGWLTAHGGTNVVASFSWNGMALN 253
Db      170 ----- 169
QY      254 PDNLIISKDVTGATKVKYYYAVNDGPGDGHYAVMISKGTNAGDFTVVFEETPNKGG 313
Db      170 -----YTP----- 172
QY      314 ARFGLSTEADGAKPQSVWIERVDLPAGTKYVAFRHYNCSDLNLYLLDDIOFTMGGSPTP 373
Db      173 -----LP-----RRAPC-----PHP 182
QY      374 TDYTYTVVRDGTIKIEGLTETTFEDGVATGNHEYCVVEVKYTAGVSPKCVNVTV 428
Db      183 GGYTYSVFRDQKIASGLSALAYIDTVPYQDICYQVQVNYLQGDYSYKVCNKIVV 237

RESULT 15
Q9XBU9
ID Q9XBU9 PRELIMINARY; PRT; 293 AA.
AC Q9XBU9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Immunoreactive 32 kDa antigen PG25.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_taxid=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RT Hocking D., Webb E.;
RL "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF155351; AAD38410.1; -.
SQ SEQUENCE 293 AA; 32272 MW; CC03EAC241F7F6F1 CRC64;

Query Match 9.4%; Score 233; DB 2; Length 293;
Best Local Similarity 59.0%; Pred. No. 1.7e-07;
Matches 46; Conservative 7; Mismatches 23; Indels 2; Gaps 1;

QY      376 YTTVVRDGTIKIEGLTETTFEDGVATGNHEYCVVEVKYTAGVSPKCVNVTVDPQFNP 435
Db      77 YTTVVRDGTIKIEGLTETTFEDGVATGNHEYCVVEVKYTAGVSPKCVNVTVDPQFNP 435
QY      436 VQNLTGSAVGQKVTLKWD 453
Db      135 VTNLGTASNDEVSLDWD 152
```

Search completed: May 18, 2004, 11:46:38
Job time : 35.3368 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 14.0888 Seconds
(without alignments)
1670.936 Million cell updates/sec

Title: US-08-570-311-18
Perfect score: 2480
Sequence: 1 GTPNPNPNPNTTLESEF.....QNLTGSAVGQVKTLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------------|
| 1 | 2480 | 100.0 | 456 | 2 | US-08-570-311-18 |
| 2 | 2480 | 100.0 | 456 | 2 | US-08-570-311-20 |
| 3 | 2480 | 100.0 | 2628 | 2 | US-08-570-311-14 |
| 4 | 2436 | 98.2 | 450 | 2 | US-08-570-311-16 |
| 5 | 2351 | 94.8 | 439 | 2 | US-08-570-311-22 |
| 6 | 2082 | 84.0 | 1477 | 4 | US-09-482-500A-1 |
| 7 | 2082 | 84.0 | 1687 | 2 | US-08-570-311-29 |
| 8 | 2082 | 84.0 | 1704 | 3 | US-08-336-308A-10 |
| 9 | 2082 | 84.0 | 1704 | 3 | US-08-822-324-6 |
| 10 | 2082 | 84.0 | 1704 | 3 | US-09-490-931-10 |
| 11 | 2070 | 83.5 | 1706 | 4 | US-09-066-330-10 |
| 12 | 2066 | 83.3 | 1087 | 2 | US-08-570-311-8 |
| 13 | 2066 | 83.3 | 1087 | 2 | US-08-353-485-8 |
| 14 | 2066 | 83.3 | 1358 | 2 | US-08-570-311-27 |
| 15 | 2039.5 | 82.2 | 1732 | 4 | US-09-066-330-11 |
| 16 | 2034.5 | 82.0 | 1732 | 2 | US-08-570-311-10 |
| 17 | 2034.5 | 82.0 | 1732 | 2 | US-08-353-485-10 |
| 18 | 546.5 | 22.0 | 497 | 2 | US-08-570-311-2 |
| 19 | 546.5 | 22.0 | 497 | 2 | US-08-353-485-2 |
| 20 | 296.5 | 12.0 | 942 | 1 | US-08-141-324-14 |
| 21 | 296.5 | 12.0 | 942 | 1 | US-08-541-902-14 |
| 22 | 236 | 9.5 | 49 | 3 | US-08-822-324-18 |
| 23 | 189 | 7.6 | 46 | 3 | US-08-822-324-9 |
| 24 | 137 | 5.5 | 1833 | 4 | US-08-621-944A-4 |
| 25 | 137 | 5.5 | 1833 | 4 | US-08-945-567D-4 |
| 26 | 137 | 5.5 | 1992 | 4 | US-08-621-944A-3 |
| 27 | 137 | 5.5 | 1992 | 4 | US-08-945-567D-3 |

28 134.5 5.4 2736 4 US-09-252-991A-30227 Sequence 30227, A
29 132.5 5.3 2315 4 US-09-543-681A-5434 Sequence 5434, A
30 131.5 5.3 492 4 US-09-482-500A-2 Sequence 2, Appli
31 131.5 5.3 737 1 US-08-119-361-5 Sequence 5, Appli
32 131.5 5.3 737 3 US-08-336-308A-4 Sequence 4, Appli
33 131.5 5.3 737 3 US-08-822-324-4 Sequence 4, Appli
34 131.5 5.3 737 3 US-09-490-931-4 Sequence 4, Appli
35 131.5 5.3 2123 3 US-08-968-685A-10 Sequence 10, Appli
36 129.5 5.2 1638 4 US-09-071-035-258 Sequence 258, App
37 129.5 5.2 1638 4 US-09-071-035-262 Sequence 262, App
38 129.5 5.2 1638 4 US-09-071-035-266 Sequence 266, App
39 129.5 5.2 1747 4 US-09-134-000C-5999 Sequence 5999, Ap
40 129 5.2 24 4 US-09-066-330-6 Sequence 6, Appli
41 128 5.2 25 3 US-08-822-324-19 Sequence 19, Appli
42 126 5.1 2411 4 US-09-268-347-36 Sequence 36, Appli
43 123.5 5.0 669 4 US-09-071-035-264 Sequence 264, App
44 123.5 5.0 2048 4 US-09-268-347-48 Sequence 48, Appli
45 121 4.9 509 3 US-08-822-324-8 Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-570-311-18
; Sequence 18, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tunwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 456 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-18

Query Match      100.0%; Score 2480; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.8e-204;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTNPENPNPGTTLTSESFENGIPASWKTIIDADGDNWNTTTPPGGTSFAGHNSAICA 60
Db 1 GTNPENPNPGTTLTSESFENGIPASWKTIIDADGDNWNTTTPPGGTSFAGHNSAICA 60
QY 61 SSASYINFEGPQNPNDYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 61 SSASYINFEGPQNPNDYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEEVLTAKTVVTAPPAIRGTRVQGTWYQKTVQKTVQKTVQKTVQKTVQKTVQKTV 180
Db 121 FANALLEEVLTAKTVVTAPPAIRGTRVQGTWYQKTVQKTVQKTVQKTVQKTVQKTVQKTV 180
QY 181 DDVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWCLSSGQGLGWLTAHGGTN 240
Db 181 DDVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWCLSSGQGLGWLTAHGGTN 240
QY 241 VVASFWSNGMALPNPNLYLISKDVGTGATKVKYVAVNDGPPGDHYAVMISKGTGNAGDFTV 300
Db 241 VVASFWSNGMALPNPNLYLISKDVGTGATKVKYVAVNDGPPGDHYAVMISKGTGNAGDFTV 300
QY 301 VFEETPNGINKGARGFLSTEADGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
Db 301 VFEETPNGINKGARGFLSTEADGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
QY 361 DDQFTMGSGPTDPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
Db 361 DDQFTMGSGPTDPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
QY 421 KECVNVTVDPVQNPVQNLTSAGVQKVTILKWDAPN 456
Db 421 KECVNVTVDPVQNPVQNLTSAGVQKVTILKWDAPN 456

RESULT 2
US-08-570-311-20
; Sequence 20, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-20

```

```

Query Match      100.0%; Score 2480; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.8e-204;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTNPENPNPGTTLTSESFENGIPASWKTIIDADGDNWNTTTPPGGTSFAGHNSAICA 60
Db 1 GTNPENPNPGTTLTSESFENGIPASWKTIIDADGDNWNTTTPPGGTSFAGHNSAICA 60
QY 61 SSASYINFEGPQNPNDYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 61 SSASYINFEGPQNPNDYLVTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEEVLTAKTVVTAPPAIRGTRVQGTWYQKTVQKTVQKTVQKTVQKTVQKTVQKTV 180
Db 121 FANALLEEVLTAKTVVTAPPAIRGTRVQGTWYQKTVQKTVQKTVQKTVQKTVQKTVQKTV 180
QY 181 DDVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWCLSSGQGLGWLTAHGGTN 240
Db 181 DDVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWCLSSGQGLGWLTAHGGTN 240
QY 241 VVASFWSNGMALPNPNLYLISKDVGTGATKVKYVAVNDGPPGDHYAVMISKGTGNAGDFTV 300
Db 241 VVASFWSNGMALPNPNLYLISKDVGTGATKVKYVAVNDGPPGDHYAVMISKGTGNAGDFTV 300
QY 301 VFEETPNGINKGARGFLSTEADGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
Db 301 VFEETPNGINKGARGFLSTEADGAKPQSWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
QY 361 DDQFTMGSGPTDPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
Db 361 DDQFTMGSGPTDPTDYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
QY 421 KECVNVTVDPVQNPVQNLTSAGVQKVTILKWDAPN 456
Db 421 KECVNVTVDPVQNPVQNLTSAGVQKVTILKWDAPN 456

```

```

RESULT 3
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn

```

```

; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-14

Query Match 100.0%; Score 2480; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 3.8e-203;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTNPENPNPGTTTISEFENGIPASWKTIADGDGNNWTTTTPPGGTSFAGHNSAICA 60
Db 950 GTNPENPNPGTTTISEFENGIPASWKTIADGDGNNWTTTTPPGGTSFAGHNSAICA 1009
QY 61 SSASYINFEGQPNPNLVTPPELSPNGGTLTFWCAQDANYASEHVAVYASSTGNDASN 120
Db 1010 SSASYINFEGQPNPNLVTPPELSPNGGTLTFWCAQDANYASEHVAVYASSTGNDASN 1069
QY 121 FANALLEVLTAKTVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAPRHFPGCTDFFWNL 180
Db 1070 FANALLEVLTAKTVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAPRHFPGCTDFFWNL 1129
QY 181 DVEIKANGRADPTETFEFSTHGEAPAEWTTTADGDGGQWCLSSGQLGWLTAHGNTN 240
Db 1130 DVEIKANGRADPTETFEFSTHGEAPAEWTTTADGDGGQWCLSSGQLGWLTAHGNTN 1189
QY 241 VVASFWSNGMALPNPNLVLSKDVTKYKYYAVNDGFPDGHVAVMISKTGTNAGDFTV 300
Db 1190 VVASFWSNGMALPNPNLVLSKDVTKYKYYAVNDGFPDGHVAVMISKTGTNAGDFTV 1249
QY 301 VFETPNNGKNGARFGLSTADGAKPQSVWIERTVLPAGTKYVAPRHYNCSDLNILL 360
Db 1250 VFETPNNGKNGARFGLSTADGAKPQSVWIERTVLPAGTKYVAPRHYNCSDLNILL 1309
```

```

QY 361 DDQFTMGSGSPTDPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEVKTAGVSP 420
Db 1310 DDQFTMGSGSPTDPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEVKTAGVSP 1369
QY 421 KECVNVTVDPVQPNPNVQNLTGSAVGQKVTILKWDAPN 456
Db 1370 KECVNVTVDPVQPNPNVQNLTGSAVGQKVTILKWDAPN 1405
```

RESULT 4

```

US-08-570-311-16
; Sequence 16, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guyline
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-16
```

```

Query Match 98.2%; Score 2436; DB 2; Length 450;
Best Local Similarity 99.8%; Pred. No. 1.6e-200;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 PNPENPGTTTISEFENGIPASWKTIADGDGNNWTTTTPPGGTSFAGHNSAICA 66
Db 1 PNPENPGTTTISEFENGIPASWKTIADGDGNNWTTTTPPGGTSFAGHNSAICA 60
```

QY 67 NFEQPNDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
Db 61 NFEQPNDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
QY 127 BEVLTAKTAVTAPPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINLDDVEIK 186
Db 121 BEVLTAKTAVTAPPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINLDDVEIK 180
QY 187 ANGKRADETFESTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTNNVASFS 246
Db 181 ANGKRADETFESTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTNNVASFS 240
QY 247 WNGMALNPDNVLISKDVTGATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFTVVFEETP 306
Db 241 WNGMALNPDNVLISKDVTGATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFTVVFEETP 300
QY 307 NGINKGGARFGLSTADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNLYILLDDIOFT 366
Db 301 NGINKGGARFGLSTADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNLYILLDDIOFT 360
QY 367 MGSSTPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVVXYTAGVSPKECVNV 426
Db 361 MGSSTPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVVXYTAGVSPKECVNV 420
QY 427 TVDPQFNPQNLGSAVGQKVTLLKWDAPN 456
Db 421 TVDPQFNPQNLGSAVGQKVTLLKWDAPN 450

RESULT 5

US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 439 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-22

Query Match 94.8%; Score 2351; DB 2; Length 439;
Best Local Similarity 98.2%; Pred. No. 3e-193;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 GTPNPNPNPNTGTTLSSEFENGIPASWKTTIDADGNGNWTTPPGGTSPAGHNSAICA 60
Db 1 GTPNPNPNPNTGTTLSSEFENGIPASWKTTIDADGNGNWTTPPGGTSPAGHNSAICA 60
QY 61 SSASYINPEGPQNPNDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 61 SSASYINPEGPQNPNDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEVLTAKTAVTAPPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINL 180
Db 121 FANALLEVLTAKTAVTAPPAIRGTRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWINL 180
QY 181 DDVEIKANGKRADETFESTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTN 240
Db 181 DDVEIKANGKRADETFESTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGHGTN 240
QY 241 VVAFSFWNGMALNPDNVLISKDVTGATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFTV 300
Db 241 VVAFSFWNGMALNPDNVLISKDVTGATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFTV 300
QY 301 VFEETPNGINKGGARFGLSTADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNLYILL 360
Db 301 VFEETPNGINKGGARFGLSTADGAKPOSVMWERTVLDLPAGTKYVAFRHYNCSDLNLYILL 360
QY 361 DDIOFTMGSPPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVVXYTAGVSP 420
Db 361 DDIOFTMGSPPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCVVXYTAGVSP 420
QY 421 KECVNVTVDPVQFNPQNL 439
Db 421 KVCVNVNTINPTQFNPQNL 439

RESULT 6

US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-482-500A-1
Query Match 84.0%; Score 2082; DB 4; Length 1477;


```

Best Local Similarity 86.2%; Pred. No. 2.1e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNTGTTTSLSEFENGIPASWKTIDADGDNNTTTPPGGTSFAGHNSAICASS 62
Db 725 PNPENPNTGTTTSLSEFENGIPASWKTIDADGDNNTTTPPGGTSFAGHNSAICASS 781

QY 63 ASY-INFEQPNPDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 782 ESFGLGGIGVLPDNYLITFALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841

QY 122 ANALLEVTAKTVTAPAIRTRVQGTWYQKTVDLPAGTKYVAFRHFGCTDFFWINLD 181
Db 842 TNALLETTITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFGCTDFFWINLD 899

QY 182 DVEIKANGKRADETFETFESETHGEAPAEWTTIDADGGQGLCLSSGQLDWTAGHGTNV 241
Db 900 EVEIKANGKRADETFETFESETHGEAPAEWTTIDADGGQGLCLSSGQLDWTAGHGTNV 959

QY 242 VASFSWNGMALNPNDNYLISKDVTGATKVKYIYAVNDGFGPDGHYAVMISKTGTNAGDFTVV 301
Db 960 VASFSWNGMALNPNDNYLISKDVTGATKVKYIYAVNDGFGPDGHYAVMISKTGTNAGDFTVV 1019

QY 302 FEETPNKGGARFGLSTADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILLD 361
Db 1020 FEETPNKGGARFGLSTADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILLD 1079

QY 362 DIQFTWGGSPPTDYTVTVYRDGKTKKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPK 421
Db 1080 DIQFTWGGSPPTDYTVTVYRDGKTKKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPK 1139

QY 422 ECNVNVTVDVQFNPVQNLTGSVQKQVTLKWDAPN 456
Db 1140 ECNVNVTINPTQFNPVKNLKQAPDGGDVVLKWEAPS 1174

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-29

Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 2.5e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNTGTTTSLSEFENGIPASWKTIDADGDNNTTTPPGGTSFAGHNSAICASS 62
Db 935 PNPENPNTGTTTSLSEFENGIPASWKTIDADGDNNTTTPPGGTSFAGHNSAICASS 991

QY 63 ASY-INFEQPNPDNYLVTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 992 ESFGLGGIGVLPDNYLITFALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051

QY 122 ANALLEVTAKTVTAPAIRTRVQGTWYQKTVDLPAGTKYVAFRHFGCTDFFWINLD 181
Db 1052 TNALLETTITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFGCTDFFWINLD 1109

QY 182 DVEIKANGKRADETFETFESETHGEAPAEWTTIDADGGQGLCLSSGQLDWTAGHGTNV 241
Db 1110 EVEIKANGKRADETFETFESETHGEAPAEWTTIDADGGQGLCLSSGQLDWTAGHGTNV 1169

QY 242 VASFSWNGMALNPNDNYLISKDVTGATKVKYIYAVNDGFGPDGHYAVMISKTGTNAGDFTVV 301
Db 1170 VASFSWNGMALNPNDNYLISKDVTGATKVKYIYAVNDGFGPDGHYAVMISKTGTNAGDFTVV 1229

QY 302 FEETPNKGGARFGLSTADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILLD 361
Db 1230 FEETPNKGGARFGLSTADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILLD 1289

QY 362 DIQFTWGGSPPTDYTVTVYRDGKTKKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPK 421
Db 1290 DIQFTWGGSPPTDYTVTVYRDGKTKKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPK 1349

QY 422 ECNVNVTVDVQFNPVQNLTGSVQKQVTLKWDAPN 456
Db 1350 ECNVNVTINPTQFNPVKNLKQAPDGGDVVLKWEAPS 1384

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado

```

```
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-308A-10

Query Match 84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.6e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSLSEFENGIPASWKITDADGDNNTTTPPGGTSPAGHNSAICASS 62
Db 952 PNPENPNPGTTLTSLSEFENGIPASWKITDADGDNNTTTPPGGTSPAGHNSAICASS 1008

QY 63 ASY-INFEQPNPDNYLVTPELSLPGNGTLTFWVCAQADANYASEHYAVYASSTGNDASNF 121
Db 1009 ESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQADANYASEHYAVYASSTGNDASNF 1068

QY 122 ANALLEEVLTAKTVTAPPAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFPGCTDFFWINLD 181
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYDLD 1126

QY 182 DVEIKANGKRAADPTTFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTVN 241
Db 1127 EVEIKANGKRAADPTTFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTVN 1186

QY 242 VASFSWNGMALNPDNYLISKDVTGATKVKYIYAVNDGPGDHYAVMISKTGTNAGDFTVV 301
Db 1187 VASFSWNGMALNPDNYLISKDVTGATKVKYIYAVNDGPGDHYAVMISKTGTNAGDFTVV 1246

QY 302 FEETPNKNGGARFGLSTEADGAKPOSVMERTVDPAGTKYVAFRHYNCSDLNILLD 361
Db 1247 FEETPNKNGGARFGLSTEADGAKPOSVMERTVDPAGTKYVAFRHYNCSDLNILLD 1306

QY 362 DIQFTMGSSPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEYCVVEKYTAGVSPK 421
Db 1307 DIQFTMGSSPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEYCVVEKYTAGVSPK 1366

QY 422 ECNVTVDPVQFPVQNLTSANGQKVTLLKWDAPN 456
Db 1367 ECNVTVINPTQFNPKVLKKAQPDGDDVVLKWEAPS 1401

RESULT 9
US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917
```

[illegible][illegible]

Db 1309 DIQFTMGSPPTDYTYTVYRDGTGKKEGLTETTFEEDGVATGNHCVKVTAGVSPK 1368

QY 422 ECNVTVDPQFNPVQNLTSAGVQKVTWKWDAPN 456

Db 1369 KCVNVTVNSTQFNPVKNLKAQPDGDDVVLKWEAPS 1403

RESULT 12

US-08-570-311-8

; Sequence 8, Application US/08570311

; Patent No. 5824791

; GENERAL INFORMATION:

; APPLICANT: Proguiske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guylaine

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; NUMBER OF INVENTION: and Probes for the Detection of Periodontal Disease

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ted W. Whitlock

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/570,311

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/353,485

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/647,119

; FILING DATE: 25-JAN-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/241,640

; FILING DATE: 08-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UF15.C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (904) 375-8100

; TELEFAX: (904) 372-5800

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1087 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-570-311-8

Query Match 83.3%; Score 2066; DB 2; Length 1087;

Best Local Similarity 85.3%; Pred. No. 3.1e-168;

Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

QY 1 GTPNPNPNPGTTTSSSFNGIPASWKTIDADGDKNNWTTTPPGTSPAGNSAICA 60

Db 333 GTPNPNPNPGTTTSSSFNGIPASWKTIDADGDKNNWTTTPPGTSPAGNSAICA 60

QY 61 SSASY-INFEQFQPDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119

Db 390 YSBSFGLGGIGVLTDPNLYLITPALDLANGCKLTFWVCAQDANYASEHYAVYASSTGNDAS 449

QY 120 NFANALLEEVLTAKTVVTAPEAIRGTVOGTWOKTVOLPAGTKYVAFRHFQCTDFWIN 119

Db 450 NFNALLEEITITAKG-VRSPEAIRG-RIQSTMRQKTVDLPAQTKYVAFRHFQCTDMFYID 507

QY 180 LDDVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGQCGWMLCLSSGQLWLTAGGT 239

Db 508 LDEVEIKANGKRADFTETFESSTHGEAPAEWTTIDADGQCGWMLCLSSGQLWLTAGGT 567

QY 240 NVVASFSWNGMALNPDNYLISKDVGTATKVKYVAVNDGPPGPDHYAVMISKTGTNAGDFT 299

Db 568 NVVASFSWNGMALNPDNYLISKDVGTATKVKYVAVNDGPPGPDHYAVMISKTGTNAGDFT 627

QY 300 VFEETPNGINKGARFGLSTEADGAKPOSVMTERVTDLPAGTKYVAFRHYNSDLYIL 359

Db 628 VFEETPNGINKGARFGLSTEADGAKPOSVMTERVTDLPAGTKYVAFRHYNSDLYIL 687

QY 360 LDDIQFTMGSPPTDYTYTVYRDGTGKKEGLTETTFEEDGVATGNHCVKVTAGVS 419

Db 688 LDDIQFTMGSPPTDYTYTVYRDGTGKKEGLTETTFEEDGVATGNHCVKVTAGVS 747

QY 420 PKECVNVTVDPQFNPVQNLTSAGVQKVTWKWDAPN 456

Db 748 PKVCNVTVINPTQFNPVKNLKAQPDGDDVVLKWEAPS 784

RESULT 13

US-08-353-485-8

; Sequence 8, Application US/08353485

; Patent No. 5830710

; GENERAL INFORMATION:

; APPLICANT: Proguiske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guylaine

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ted W. Whitlock

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/353,485

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/647,119

; FILING DATE: 25-JAN-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/241,640

; FILING DATE: 08-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UF15.C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (904) 375-8100

; TELEFAX: (904) 372-5800

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:


```
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match      82.2%; Score 2039.5; DB 4; Length 1732;
Best Local Similarity 84.5%; Pred. No. 1.2e-165;
Matches 388; Conservative 18; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPENPNPGT-TTILSEFENGIPASWKTTIDADGDNWTTTPPGCTSPAGHNSAIC 59
Db 969 GTPNPENPNPGTTLSEFENGIPASWKTTIDADGDNWTTTPPGCTSPAGHNSAIC 1025

QY 60 ASSASY-INFGPQNPNDVLPESLPNGGTLTFWCAQDANYASEHYAVYASSTGDA 118
Db 1026 VYSEFGLGGIGVLTFDNLITPALDLPNGGKLTFWCAQDANYASEHYAVYASSTGDA 1085

QY 119 SNFANALLEEVLTAKTVTTPAIRAIGTRVQGTWYQKTVOLPAGTKYVAFRHFEGCTDFWI 178
Db 1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYI 1143

QY 179 NLDDVEIKANGRADFTETFESETHGEAPAEWTTIDADGGQGLCLSSGQLDLTAHGG 238
Db 1144 DLDEVEIKANGRADFTETFESETHGEAPAEWTTIDADGGQGLCLSSGQLDLTAHGG 1203

QY 239 TNVVASFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGNAGDF 298
Db 1204 SNVVSFSWNGMALNPDNYLISKDVTGATKVKYYAVNDGPPGDHYAVMISKTGNAGDF 1263

QY 299 TVVFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLDPAGTKYVAFRHYNCSDLNYI 358
Db 1264 TVVFEETPNGINKGARFGLSTEANGAKPQSVWIERTVDLDPAGTKYVAFRHYNCSDLNYI 1323

QY 359 LLDDIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCUEVKYTAGV 418
Db 1324 LLDDIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCUEVKYTAGV 1383

QY 419 SPKCVNVTVPVQNPVNLTGSAV--GQVTLKMDAP 455
Db 1384 SPKCVNVTVNSTQNPVNLTAEQAPNSMDAILKKNAP 1422
```

Search completed: May 18, 2004, 11:49:14
Job time : 15.0888 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:42:50 ; Search time 35.3211 Seconds
(without alignments)

3592.387 Million cell updates/sec

Title: US-08-570-311-18

Perfect score: 2480

Sequence: 1 GTPNPENPNPGTTLTSESF.....QNLTSAGVCKVTLKWDAPN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 2070 | 83.5 | 1706 | 14 | US-10-229-066-10 |
| 2 | 2039.5 | 82.2 | 1732 | 14 | US-10-229-066-11 |
| 3 | 726 | 29.3 | 419 | 15 | US-10-174-695-3 |
| 4 | 710.5 | 28.6 | 419 | 15 | US-10-174-695-5 |
| 5 | 385 | 15.5 | 231 | 15 | US-10-174-695-6 |
| 6 | 232.5 | 9.4 | 196 | 15 | US-10-174-695-4 |
| 7 | 153 | 6.2 | 29 | 15 | US-10-387-977-18 |
| 8 | 153 | 6.2 | 29 | 15 | US-10-387-977-21 |
| 9 | 153 | 6.2 | 29 | 15 | US-10-387-977-23 |
| 10 | 153 | 6.2 | 29 | 15 | US-10-387-977-24 |
| 11 | 153 | 6.2 | 29 | 15 | US-10-387-977-26 |
| 12 | 147 | 5.9 | 29 | 15 | US-10-387-977-17 |
| 13 | 147 | 5.9 | 872 | 12 | US-10-282-122A-55467 |
| 14 | 145 | 5.8 | 29 | 15 | US-10-387-977-25 |
| 15 | 143 | 5.8 | 509 | 15 | US-10-387-977-101 |

| | | | | | | |
|----|-------|-----|------|----|----------------------|--------------------|
| 16 | 137 | 5.5 | 1833 | 14 | US-10-175-282-4 | Sequence 4, Appli |
| 17 | 137 | 5.5 | 1833 | 14 | US-10-175-275-4 | Sequence 4, Appli |
| 18 | 137 | 5.5 | 1992 | 14 | US-10-175-282-3 | Sequence 3, Appli |
| 19 | 137 | 5.5 | 1992 | 14 | US-10-175-275-3 | Sequence 3, Appli |
| 20 | 136 | 5.5 | 29 | 15 | US-10-387-977-20 | Sequence 20, Appli |
| 21 | 134 | 5.4 | 25 | 15 | US-10-387-977-77 | Sequence 77, Appli |
| 22 | 134 | 5.4 | 25 | 15 | US-10-387-977-80 | Sequence 80, Appli |
| 23 | 134 | 5.4 | 27 | 15 | US-10-387-977-11 | Sequence 11, Appli |
| 24 | 134 | 5.4 | 27 | 15 | US-10-387-977-12 | Sequence 12, Appli |
| 25 | 134 | 5.4 | 27 | 15 | US-10-387-977-13 | Sequence 13, Appli |
| 26 | 132.5 | 5.3 | 555 | 14 | US-10-156-761-14592 | Sequence 14592, A |
| 27 | 132.5 | 5.3 | 691 | 15 | US-10-369-493-23473 | Sequence 23473, A |
| 28 | 131.5 | 5.3 | 2122 | 9 | US-09-813-214A-9 | Sequence 9, Appli |
| 29 | 131.5 | 5.3 | 2468 | 12 | US-10-282-122A-66335 | Sequence 66335, A |
| 30 | 131.5 | 5.3 | 2468 | 14 | US-10-246-330-4 | Sequence 4, Appli |
| 31 | 131 | 5.3 | 960 | 14 | US-10-311-879-15 | Sequence 15, Appli |
| 32 | 130 | 5.2 | 25 | 15 | US-10-387-977-81 | Sequence 81, Appli |
| 33 | 130 | 5.2 | 1946 | 12 | US-10-282-122A-62947 | Sequence 62947, A |
| 34 | 129.5 | 5.2 | 1638 | 12 | US-10-206-576-258 | Sequence 258, App |
| 35 | 129.5 | 5.2 | 1638 | 12 | US-10-206-576-262 | Sequence 262, App |
| 36 | 129.5 | 5.2 | 1638 | 12 | US-10-206-576-266 | Sequence 266, App |
| 37 | 129.5 | 5.2 | 1728 | 12 | US-10-282-122A-56997 | Sequence 56997, A |
| 38 | 129 | 5.2 | 24 | 14 | US-10-229-066-6 | Sequence 6, Appli |
| 39 | 129 | 5.2 | 2435 | 12 | US-10-282-122A-47453 | Sequence 47453, A |
| 40 | 127 | 5.1 | 25 | 15 | US-10-387-977-83 | Sequence 83, Appli |
| 41 | 127 | 5.1 | 25 | 15 | US-10-387-977-84 | Sequence 84, Appli |
| 42 | 127 | 5.1 | 25 | 15 | US-10-387-977-85 | Sequence 85, Appli |
| 43 | 126 | 5.1 | 1649 | 15 | US-10-369-493-18460 | Sequence 18460, A |
| 44 | 123.5 | 5.0 | 491 | 15 | US-10-387-977-100 | Sequence 100, App |
| 45 | 123.5 | 5.0 | 669 | 12 | US-10-206-576-264 | Sequence 264, App |

ALIGNMENTS

RESULT 1

US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Shakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 83.5%; Score 2070; DB 14; Length 1706;

Best Local Similarity 85.7%; Pred. No. 3.9e-176;

Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTTIDAGDGNNTTTPPGGTSEFAGHNSAICASS 62

Db 954 PNPENPNPGTTLTSESFENGIPASWKTTIDAGDGHGKPGNAPG---IAGYNSGCVYS 1010

QY 63 ASY-IFNFGPONPDNLYVTPELSLFPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121

Db 1011 BSFGLGGIGVLTPTDNYLTLPALDLENGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

| | | | |
|----|------|--|------|
| Qy | 122 | ANALLEVLTAKTVVTAPAEIRGRTRVQGTWYQKTVQLPAGTKYVAFRHFPGCTDFWINLD | 181 |
| Db | 1071 | TNALLEETITAKG-VRSEAPMRG-RIOGTWRQKTVDLFAGTKYVAFRHFQSTDMFYDLD | 1128 |
| Qy | 182 | DVEIKANGKRAADFETTESSTHGAPAEWTTIDAGDQGQWCLSSGQLGWLTAGGTNV | 241 |
| Db | 1129 | EVEIKANGKRAADFETTESSTHGAPAEWTTIDAGDQGQWCLSSGQLDWLTAGGTNV | 1188 |
| Qy | 242 | VASFSGWGMALPNPNYLISKDVTGATKVKYKYAYVNDGFPGDHYAVMLSKGTNAGDFTVV | 301 |
| Db | 1189 | VSSFSGWGMALPNPNYLISKDVTGATKVKYKYAYVNDGFPGDHYAVMLSKGTNAGDFTVV | 1248 |
| Qy | 302 | FEETPNGINKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLD | 361 |
| Db | 1249 | FEETPNGINKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLD | 1308 |
| Qy | 362 | DIQFTMGSPPTDYTTVVRDGTGKKEGTTEITFEEDGATGNHGYCBEVKYTAGVSPK | 421 |
| Db | 1309 | DIQFTMGSPPTDYTTVVRDGTGKKEGTTEITFEEDGATGNHGYCBEVKYTAGVSPK | 1368 |
| Qy | 422 | ECNVTVDPQVQFNPQNLTSAGVQKVTLKWDAPN | 456 |
| Db | 1369 | KCNVTVNSTQFNPVKNLKXQPDGDEGVVLKWEAPS | 1403 |

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

| | | | | | | | | | |
|-----------------------|-------|--------------|----------|------------|--------|--------|---|------|---|
| Query Match | 82.2% | Score | 2039.5 | DB 14 | Length | 1732 | | | |
| Best Local Similarity | 84.5% | Pred. No. | 2.2e-173 | | | | | | |
| Matches | 398 | Conservative | 18 | Mismatches | 44 | Indels | 9 | Gaps | 6 |

| | | | |
|----|------|---|------|
| QY | 1 | GTNPENPNPGT-TTSSSFENGISASWKTIIDADGGNNWTTTPPPGGTSPAGHNSAIC | 59 |
| DB | 969 | GTNPENPNPNPGTTLSSSFENGISASWKTIIDADGGHGWKPGNAPG---IAGYNSGC | 1025 |
| | | | |
| QY | 60 | ASSASY-INPEGPQNPDNVLTPELSPNGGTLTPWVCAQDANYASEHVAVASSTGND | 118 |
| DB | 1026 | VYSESEGLGIGVLTDPDNYLTIPALDIPNGGKLTTPWVCAQDANYASEHVAVASSTGND | 1085 |
| | | | |
| QY | 119 | SNFANALLBEVLTAKTIVTAPPAIRTRVOGTWYQTVOLPAGTKYVAFRHFGCTDFPW | 178 |
| DB | 1086 | SNFTNALLBETITAKG-VRSPKAIRG-RIQGTWRQKTVOLPAGTKYVAFRHFGQSTDMFYI | 1143 |
| | | | |
| QY | 179 | NLDDEVIKANGKREADFTETFESSTHGEAPAEWTTIIDADGGQGWCLSSSQGLMHTAGG | 238 |
| DB | 1144 | DLDEVEIKANGKREADFTETFESSTHGEAPAEWTTIIDADGGQGWCLSSSQGLDMLTAGG | 1203 |
| | | | |
| QY | 239 | TNVVASFSWNGMALPNPNYLI SKDVTGATKVKYKYAVNDGPGGDHVAVMISKGTGNAGDF | 298 |
| | | | |

| | | | |
|----|------|---|------|
| Db | 1204 | SNVSSFSMNGMALNPNDNYILSKDVTGATKVXYAVNDGPGDHYAVMISKTGTNAGDF | 1263 |
| Qy | 299 | TVVFEETPNGINGKGGARFGLSTBADGAKPOSWIERVDLPACTKYVAPRHYNCSOLNYI | 358 |
| Db | 1264 | TVVFEETPNGINGKGGARFGLSTBANGAKPOSWIERVDLPACTKYVAFRHYNCSOLNYI | 1323 |
| Qy | 359 | LLDDIOFTMGSGPTPTDYTVTVYRDGTGKEGILTETTFEEDGVATGNHGYCUEVKVKTAGV | 418 |
| Db | 1324 | LLDDIOFTMGSGPTPTDYTVTVYRDGTGKEGILTETTFEEDGVATGNHGYCUEVKVKTAGV | 1383 |
| Qy | 419 | SPKECVNVTVDPVQFNPVQNLGTSV--GOKVTLKWDAP | 455 |
| Db | 1384 | SPKKCVNVTVNSQFNPVQNLTAEQAPNMSDAILTKWNAP | 1422 |

```

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US2003023202A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

```

| | | | | | | | |
|-----------------------|--------|--------------|----------|------------|-----|--------|------|
| Query Match | 29.3%; | Score | 726; | DB | 15; | Length | 419; |
| Best Local Similarity | 75.3%; | Pred. No. | 2.3e-56; | | | | |
| Matches | 143; | Conservative | 11; | Mismatches | 30; | Indels | 6; |
| | | | | | | Gaps | 4; |

| | | | |
|----|-----|---|-----|
| QY | 3 | PNPNPNPQTTLLSSSFENGIPASAKNTIDADGGNNWTTTPPGGTSFACHNSAICASS | 62 |
| | | | |
| Db | 235 | PNPNPNPQTTLLSSSFENGIPASAKNTIDADGGHGWKPCNAEG--- | 291 |
| | | | |
| QY | 63 | ASY-INFEQPNPDNYLVTEPLSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF | 121 |
| | | | |
| Db | 292 | ESFGLGGIGVLTDPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF | 351 |
| | | | |
| QY | 122 | ANALLEVLTAKTVVTAPBAIRCTRVOGTWYQKTVLPAGTKYVAFRHFPGCTDFFWNL | 181 |
| | | | |
| Db | 352 | TNALLEETITAKG-VRSPEAIRG-RIGQTRQKTVLDPAGTKYVAFRHFPOSTDMFYIDL | 409 |
| | | | |
| QY | 182 | DVEIKANGKR | 191 |
| | | | |
| Db | 410 | EVEIKANGKR | 419 |
| | | | |

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Revtoldas, Eric Charles
; APPLICANT: Slateski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174-695


```

; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5

Query Match      28.6%; Score 710.5; DB 15; Length 419;
Best Local Similarity 73.6%; Pred. No. 5.6e-55;
Matches 142; Conservative 12; Mismatches 32; Indels 7; Gaps 5;

QY 1 GTPNPNPNPGT--TTLSSEFENGIPASWKTIADGDGNNWTTTPPGTSTFAGHNSAIC 59
DB 232 GTPNPNPNPNPGTTLSEFENGIPASWKTIADGDGNNWTTTPPGTSTFAGHNSAIC 288
QY 60 ASSASY-INFEQPNPNPNYLTPELSPLNGGTLTFWVCAQDANYASEHVAVSSTGND 118
DB 289 VYSEFGLGIGVLTDPNLYLTPELSPLNGGTLTFWVCAQDANYASEHVAVSSTGND 348
QY 119 SNFANALLEEVLTAKTVTTAPEAIRGTRVGTVQKTVOLPAGTKYKVAFRHFGCTDFWI 178
DB 349 SNETNALLEETITAKG-VRSKPAIRG-RIOGTWRQKTVLDPAGTKYKVAFRHFGCTDFWI 406
QY 179 NLDDVEIKANGKR 191
DB 407 DLDEVEIKANGKR 419

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US2003023202A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4

Query Match      9.4%; Score 232.5; DB 15; Length 196;
Best Local Similarity 39.1%; Pred. No. 1.5e-12;
Matches 61; Conservative 13; Mismatches 63; Indels 19; Gaps 5;

QY 3 PNPNNPNPGTTLSEFENGIPASWKTIADGDGNNWTTTPPGTSTFAGHNSAICASS 62
DB 44 PNPNNPNPNPGTTLSEFENGIPASWKTIADGDGNNWTTTPPGTSTFAGHNSAICASS 100
QY 63 ASYINFEQPNPNPNYLTPELSPLNGGTLTFWVCAQDANYASEHVAVSSTGNDASNFA 122
DB 101 DNSAKIDRNOEINVY-NTAEYAKTNNAPI-----KVVGADE-----KTGTAAYNM- 145
QY 123 NALLEEVLTAKTVTTAPEAIRGTRVGTVQKTVOL 158
DB 146 -KLSERRAKAVAKMLEKYGVSAADRTIETWKSSSQI 180

RESULT 7
US-10-387-977-18
; Sequence 18, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PQ 6528
; PRIOR FILING DATE: 1997-04-30

```

```
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match          6.2%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
      |||||
Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 8
US-10-387-977-21
; Sequence 21, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match          6.2%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
      |||||
Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 9
US-10-387-977-23
; Sequence 23, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match          6.2%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
      |||||
Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 10
US-10-387-977-24
; Sequence 24, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match          6.2%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
      |||||
Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 11
US-10-387-977-26
; Sequence 26, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; FILE REFERENCE: 52928200301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match          6.2%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
      |||||
Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29
```

;; PRIOR APPLICATION NUMBER: AU PO 6528
;; PRIOR FILING DATE: 1997-04-30
;; NUMBER OF SEQ ID NOS: 105
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 26
;; LENGTH: 29
;; TYPE: PRT
;; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match 6.2%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DTYTYVRDGTIKKGLTETTFEEDGVAT 403
Db 1 DTYTYVRDGTIKKGLTETTFEEDGVAT 29

RESULT 12

US-10-387-977-17
;; Sequence 17, Application US/10387977
;; Publication No. US20040005276A1
;; GENERAL INFORMATION:
;; APPLICANT: Reynolds, Eric Charles
;; APPLICANT: O'Brien-Simpson, Neil Martin
;; APPLICANT: Slakeski, Nada
;; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
;; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
;; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
;; FILE REFERENCE: 529282000301
;; CURRENT APPLICATION NUMBER: US/10/387,977
;; CURRENT FILING DATE: 2003-07-18
;; PRIOR APPLICATION NUMBER: US 09/423,056
;; PRIOR FILING DATE: 2000-03-22
;; PRIOR APPLICATION NUMBER: PCT/AU98/00311
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: AU PO 6528
;; PRIOR FILING DATE: 1997-04-30
;; NUMBER OF SEQ ID NOS: 105
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 17
;; LENGTH: 29
;; TYPE: PRT
;; ORGANISM: Porphyromonas gingivalis
US-10-387-977-17

Query Match 5.9%; Score 147; DB 15; Length 29;
Best Local Similarity 96.6%; Pred. No. 4.3e-06;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 375 DTYTYVRDGTIKKGLTETTFEEDGVAT 403
Db 1 DTYTYVRDGTIKKGLTATTFEEDGVAT 29

RESULT 13

US-10-282-122A-55467
;; Sequence 55467, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangsu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.

;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 55467
;; LENGTH: 872
;; TYPE: PRT
;; ORGANISM: Enterobacter cloacae
US-10-282-122A-55467

Query Match 5.9%; Score 147; DB 12; Length 872;
Best Local Similarity 21.3%; Pred. No. 0.00064;
Matches 112; Conservative 56; Mismatches 205; Indels 154; Gaps 23;

QY 10 NFGTTTLLSEFENGIPASWKTIADGDGNNWTTTPPPGGTSPAGHNSAICASSASVINFE 69
Db 293 NTGTGTHNVTDVTLGPS--VAFNAISDDNVLNAVEKQDLSVSGTSSANLAEGTVTVTLN 350
QY 70 GPQNPNDVLTPELSPNGGTLTFWVCAQDANYASEHYAVVASST---GNDASNFANALL 126
Db 351 G-----KNYAATTAAD---GTWSLTVPAAADLAGLDHYTLSATATNGVGSVNTANLLV 402
QY 127 BEVLTAKT-----VVTAPBAIRGTRVQGTWYQKTVQLPAGTKYVAFRHEGCTDFFWI 178
Db 403 DTALETTVINTVAGDNVINAEEVAAQTISG-----KVANAEEAG----- 441
QY 179 NLDDVEIKANGKGRADFTTFESSHGEAPAEWTTTIDADGGQGMWCLSSGQGLWLTAGHG 238
Db 442 NTVTVTIGGN---SYTATVQSDLTWSVNPESVLTALGNGD--LTVSATV---TNGHGN 492
QY 239 T-----NVVASFSWNGMALNPDNVLISKDVTGATKVKYKYAVNODGPPGDHYAVMISKTGT 293
Db 493 TGAGEREIVIDASLPGLRIDT---VAGDDV-----INSTEHQNLIVTSSDGL 538
QY 294 NAGDPTVVFEETPNGINKGARFGLSTEADGAKPQSVMIERTVDLPACT----- 342
Db 539 AAG-----TTLTVTVNGKTYAASVLADGTWSAAIPADVDGALAAAGTVTVTAAGSAA 590
QY 343 -KYVAFRHYNCSDLNYILL-----DDI-----QFTMGGSPTPTD----- 375
Db 591 GNPVTISHDVTLDSAVAISDAIATDDVINAEKADLVLSGTSNVVEENQTVTITSGG 650
QY 376 --YTYTVYRDG-----TKIKEG-----LTETTFEEDGVATGNHEYCVVEKYTAGV 418
Db 651 KTYTAKVDADGNWTATVPSADLAGLKDGDASVQSVTNHAGNSASAGREYSVD-----AT 705
QY 419 SPKECVNVTVDPVQFNPNVQNLTGSA-----VGOKVTLKWD 453
Db 706 AP-----TVTIITVAGDNVINASAAAGVAISGTTTAEVGTQTVTVTLTD 748

[illegible]

Search completed: May 18, 2004, 11:52:20
Job time : 35.3211 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:33:39 ; Search time 47.0287 Seconds
(without alignments)
2739.638 Million cell updates/sec

Title: US-08-570-311-20

Perfect score: 2480

Sequence: 1 GTPNPNNPNNPNTTTLSESF.....QLTGSAGVQKVTLLKWDAPN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|--------------------|
| 1 | 2480 | 100.0 | 456 | 2 | AAR96023 P. gingiv |
| 2 | 2480 | 100.0 | 456 | 2 | AAR96022 P. gingiv |
| 3 | 2480 | 100.0 | 456 | 2 | AAR96021 P. gingiv |
| 4 | 2480 | 100.0 | 456 | 2 | AAR96020 P. gingiv |
| 5 | 2480 | 100.0 | 2628 | 2 | AAR96019 P. gingiv |
| 6 | 2480 | 100.0 | 2628 | 2 | AAR96018 P. gingiv |
| 7 | 2436 | 98.2 | 450 | 2 | AAR96017 P. gingiv |
| 8 | 2436 | 98.2 | 450 | 2 | AAR96016 P. gingiv |
| 9 | 2351 | 94.8 | 439 | 2 | AAR96015 P. gingiv |
| 10 | 2351 | 94.8 | 439 | 2 | AAR96014 P. gingiv |
| 11 | 2082 | 84.0 | 1687 | 2 | AAR96013 P. gingiv |
| 12 | 2082 | 84.0 | 1687 | 2 | AAR96012 P. gingiv |
| 13 | 2082 | 84.0 | 1704 | 2 | AAR96011 P. gingiv |
| 14 | 2082 | 84.0 | 1704 | 2 | AAR96010 P. gingiv |
| 15 | 2082 | 84.0 | 1704 | 3 | AAR96009 P. gingiv |
| 16 | 2082 | 84.0 | 1704 | 3 | AAR96008 P. gingiv |
| 17 | 2070 | 83.5 | 1706 | 2 | AAR96007 P. gingiv |
| 18 | 2066 | 83.3 | 1087 | 2 | AAR96006 P. gingiv |
| 19 | 2066 | 83.3 | 1087 | 2 | AAR96005 P. gingiv |
| 20 | 2066 | 83.3 | 1358 | 2 | AAR96004 P. gingiv |
| 21 | 2066 | 83.3 | 1358 | 2 | AAR96003 P. gingiv |
| 22 | 2034.5 | 82.0 | 1732 | 2 | AAR96002 P. gingiv |
| 23 | 2034.5 | 82.0 | 1732 | 2 | AAR96001 P. gingiv |
| 24 | 2034.5 | 82.0 | 1732 | 2 | AAR96000 P. gingiv |
| 25 | 726 | 29.3 | 419 | 4 | AAR96023 P. gingiv |

| | | | | | | |
|----|-------|------|-----|---|----------|-----------|
| 26 | 720 | 29.0 | 135 | 6 | ABP55081 | Porphyrom |
| 27 | 710.5 | 28.6 | 419 | 4 | AAU03574 | P. gingiv |
| 28 | 704 | 28.4 | 134 | 4 | ABP49217 | Peptide u |
| 29 | 682 | 27.5 | 921 | 2 | AAI34522 | Porphyrom |
| 30 | 682 | 27.5 | 922 | 2 | AAI34521 | Porphyrom |
| 31 | 682 | 27.5 | 925 | 2 | AAI34520 | Porphyrom |
| 32 | 682 | 27.5 | 938 | 2 | AAI34392 | Porphyrom |
| 33 | 565.5 | 22.8 | 377 | 2 | AAI34359 | Porphyrom |
| 34 | 561.5 | 22.6 | 312 | 2 | AAI34484 | Porphyrom |
| 35 | 546.5 | 22.0 | 497 | 2 | AAR96025 | P. gingiv |
| 36 | 546.5 | 22.0 | 497 | 2 | AAR96025 | P. gingiv |
| 37 | 415 | 16.7 | 970 | 2 | AAW69483 | Haemagglu |
| 38 | 385 | 15.5 | 231 | 4 | AAU03575 | P. gingiv |
| 39 | 267 | 10.8 | 991 | 2 | AAU03575 | P. gingiv |
| 40 | 236 | 9.5 | 49 | 2 | AAW34805 | Arg-speci |
| 41 | 233 | 9.4 | 293 | 2 | AAI34483 | Porphyrom |
| 42 | 233 | 9.4 | 299 | 2 | AAI34358 | Porphyrom |
| 43 | 232.5 | 9.4 | 196 | 4 | AAU03573 | P. gingiv |
| 44 | 189 | 7.6 | 46 | 2 | AAW34798 | Arg-speci |
| 45 | 153 | 6.2 | 29 | 2 | AAW83085 | Peptide f |

ALIGNMENTS

RESULT 1
AAR96023
ID AAR96023 standard; protein; 456 AA.
XX
AC AAR96023;
XX
XX
DT 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
DE P. gingivalis haemagglutinin hgaA Harep3 product.
XX
KW Haemagglutinin; hga; periodontal disease; vaccine; antibody; Harep3.
XX
OS Porphyromonas gingivalis; strain 381.
XX
PN WO9617936-A2.
XX
PD 13-JUN-1996.
XX
PF 11-DEC-1995; 95WO-US016108.
XX
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
XX
FA (UABR-) UAB RES FOUND.
XX
PI Proglutase-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
XX WPI; 1996-287181/29.
DR N-ESDB; AAT30647.
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
vaccination against periodontal disease.
XX
PS Claim 4; Page 110-112; 153pp; English.
XX
CC Harep3 (AAR96023) is the product of the Harep3 repeat unit (AAT30647) of
the hgaA gene (AAT30654) of P. gingivalis 318. It forms part of
haemagglutinin hgaA (see also AAR96030). Harep3 and other hgaA repeat
unit products (see also AAR96021-22 and AAR96024) can be obd. from
transformed host cells and used as vaccines to protect humans or animals
against periodontal disease. Expression in Salmonella cells allows prodn.
of live vaccine. Harep1-4 can also be used to detect the presence of anti
-P. gingivalis antibodies and to raise monoclonal antibodies for
diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 456 AA;

Query Match 100.0%; Score 2480; DB 2; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.5e-191;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTTLLSEFNGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 60
 DB 1 GTPNPNPNPGTTTLLSEFNGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 60

QY 61 SSASYINFEQPNPDNYLVTPELSPNGGTLTFWCAQDANYASEHYAVASSTGNDASN 120
 DB 61 SSASTINFEQPNPDNYLVTPELSPNGGTLTFWCAQDANYASEHYAVASSTGNDASN 120

QY 121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDFFWNL 180
 DB 121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDFFWNL 180

QY 181 DDVEIKANGKRAADTFETFSSTHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGTTN 240
 DB 181 DDVEIKANGKRAADTFETFSSTHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGTTN 240

QY 241 VVASFWSNGMALNPDNYLISKDVTGATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDFTV 300
 DB 241 VVASFWSNGMALNPDNYLISKDVTGATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDFTV 300

QY 301 VFETPNNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNILL 360
 DB 301 VFETPNNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNILL 360

QY 361 DDIOFTMGSSPTDPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
 DB 361 DDIOFTMGSSPTDPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420

QY 421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTWKWDAPN 456
 DB 421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTWKWDAPN 456

RESULT 2
 AAR96022
 ID AAR96022 standard; protein; 456 AA.
 AC AAR96022;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hAgA Harep2 product.
 XX
 KW Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harep2.
 XX
 OS Porphyromonas gingivalis; strain 381.
 XX
 PN WO9617936-A2.
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 PR 09-DEC-1994; 94US-00353485.
 XX
 PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 PI Progulsk-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX
 DR WPI: 1996-287181/29.
 DR N-PSDB; AAT30646.
 XX
 PT Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 PS Claim 4; Page 107-108; 153pp; English.
 XX

CC Harep2 (AAR96022) is the product of the Harep2 repeat unit (AAT30646) of
 CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
 CC haemagglutinin hAgA (see also AAR96030). Harep2 and other hAgA repeat
 CC unit products (see also AAR96021 and AAR96023-24) can be obt'd. from
 CC transformed host cells and used as vaccines to protect humans or animals
 CC against periodontal disease. Expression in Salmonella cells allows prodn.
 CC of live vaccine. Harep1-4 can also be used to detect the presence of anti
 CC -P. gingivalis antibodies and to raise monoclonal antibodies for
 CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX

SQ Sequence 456 AA;
 Query Match 100.0%; Score 2480; DB 2; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.5e-191;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTTLLSEFNGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 60
 DB 1 GTPNPNPNPGTTTLLSEFNGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 60

QY 61 SSASYINFEQPNPDNYLVTPELSPNGGTLTFWCAQDANYASEHYAVASSTGNDASN 120
 DB 61 SSASTINFEQPNPDNYLVTPELSPNGGTLTFWCAQDANYASEHYAVASSTGNDASN 120

QY 121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDFFWNL 180
 DB 121 FANALLEEVLTAKTVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFCTDFFWNL 180

QY 181 DDVEIKANGKRAADTFETFSSTHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGTTN 240
 DB 181 DDVEIKANGKRAADTFETFSSTHGEAPAEWTTIDADGGQGWCLSSGQLGWLTAHGTTN 240

QY 241 VVASFWSNGMALNPDNYLISKDVTGATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDFTV 300
 DB 241 VVASFWSNGMALNPDNYLISKDVTGATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDFTV 300

QY 301 VFETPNNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNILL 360
 DB 301 VFETPNNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNILL 360

QY 361 DDIOFTMGSSPTDPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
 DB 361 DDIOFTMGSSPTDPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420

QY 421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTWKWDAPN 456
 DB 421 KECVNVTVDPVQFNPVQNLTGSAVGQKVTWKWDAPN 456

RESULT 3
 AAW69491
 ID AAW69491 standard; protein; 456 AA.
 XX
 AC AAW69491;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hAgA, Harep3.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hAgA.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN US5824791-A.
 XX
 PD 20-OCT-1998.
 XX
 PF 11-DEC-1995; 95US-00570311.
 XX
 PR 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX

AC AAR96030;
 XX 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hgaA.
 XX
 KW Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.
 XX
 OS Porphyromonas gingivalis; strain 381.
 XX
 FH Key Location/Qualifiers
 FT Peptide 5..21
 FT /label= Sig_peptide
 XX
 PN WO9617936-A2.
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 XX 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX WPI; 1996-287181/29.
 DR N-PSDB; AAT30654.
 XX
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 6; Page 93-101; 153pp; English.
 XX
 CC P. gingivalis 381 haemagglutinin hgaA (AAR96030) was identified as the
 CC product of the hga gene (AAR30654) isolated as an EcoRV fragment of
 CC genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
 CC 24), can be obtained from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows production of a live vaccine. The haemagglutinin can
 CC also be used to detect the presence of anti-P. gingivalis antibodies and
 CC to raise monoclonal antibodies for diagnostic application. (Updated on 16-OCT-
 CC 2003 to standardise OS field)
 XX
 SQ Sequence 2628 AA;

Query Match 100.0%; Score 2480; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 9e-190;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTLTSESFENGIPASWKTIADGDNNTTTPPGGTGFAGHNSAICA 60
 DB 950 GTPNPNPNPGTTLTSESFENGIPASWKTIADGDNNTTTPPGGTGFAGHNSAICA 1009

QY 61 SSASYINFEQPNDNYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 DB 1010 SSASYINFEQPNDNYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

QY 121 FANALLEEVLTAKTAVTAPAIRTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 180
 DB 1070 FANALLEEVLTAKTAVTAPAIRTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 1129

QY 181 DDVEIKANGKRAADTFETSTHGEAPAEWTTIDADGGQWCLSSGQGLWLTAGGTTN 240
 DB 1130 DDVEIKANGKRAADTFETSTHGEAPAEWTTIDADGGQWCLSSGQGLWLTAGGTTN 1189

QY 241 VVASFSWNGMALPNPNLYLSKDVGTATKVKYVAVNDPFGPDHYAVMISKTGNAGDFTV 300
 DB 1190 VVASFSWNGMALPNPNLYLSKDVGTATKVKYVAVNDPFGPDHYAVMISKTGNAGDFTV 1249

QY 301 VFEEETPNKGGARFGLSTEADGAKPQSVWIERTVLPAGTKYVAFRHYNCSDLNYILL 360

DB 1250 VFEEETPNKGGARFGLSTEADGAKPQSVWIERTVLPAGTKYVAFRHYNCSDLNYILL 1309
 QY 361 DDIOFTMGSSPTPTDYTYVYRDGTKEGLTETTFEEDGVATGNHEYCVVEKVTAGVSP 420
 DB 1310 DDIOFTMGSSPTPTDYTYVYRDGTKEGLTETTFEEDGVATGNHEYCVVEKVTAGVSP 1369
 QY 421 KECNVNVTDPVQFNPNQNLGSAVGQKVTWKWDAPN 456
 DB 1370 KECNVNVTDPVQFNPNQNLGSAVGQKVTWKWDAPN 1405

RESULT 6
 AAW69488
 ID AAW69488 standard; protein; 2628 AA.
 XX
 AC AAW69488;
 XX
 DT 22-DEC-1998 (first entry)
 XX
 DE Haemagglutinin protein hgaA.
 XX
 KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN US5824791-A.
 XX
 PD 20-OCT-1998.
 XX
 PF 11-DEC-1995; 95US-00570311.
 XX
 PR 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
 XX WPI; 1998-582627/49.
 DR N-PSDB; AAV58875.
 XX
 PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease polypeptide(s).
 XX
 XX Claim 1; Col 91-110; 101pp; English.
 XX
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 SQ Sequence 2628 AA;

Query Match 100.0%; Score 2480; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 9e-190;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTLTSESFENGIPASWKTIADGDNNTTTPPGGTGFAGHNSAICA 60
 DB 950 GTPNPNPNPGTTLTSESFENGIPASWKTIADGDNNTTTPPGGTGFAGHNSAICA 1009

QY 61 SSASYINFEQPNDNYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
 DB 1010 SSASYINFEQPNDNYLTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1069

QY 121 FANALLEEVLTAKTAVTAPAIRTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 180
 DB 1070 FANALLEEVLTAKTAVTAPAIRTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 1129

QY 181 DVEIKANGKADFTTFESSTHGEAPAEWTTTIDADGGQWCLSSGQWLTARHGTN 240
 Db 1130 DVEIKANGKADFTTFESSTHGEAPAEWTTTIDADGGQWCLSSGQWLTARHGTN 1189
 QY 241 VVASFNGMALPNPNYLSKDVGTGKVKYVAVNDGPGDHYAVMISKTGNAGDFTV 300
 Db 1190 VVASFNGMALPNPNYLSKDVGTGKVKYVAVNDGPGDHYAVMISKTGNAGDFTV 1249
 QY 301 VFETENGKKGARFGLSTEADGAKPQSVWERTVDLPAGTKYVAFRHYNCSDLYNILL 360
 Db 1250 VFETENGKKGARFGLSTEADGAKPQSVWERTVDLPAGTKYVAFRHYNCSDLYNILL 1309
 QY 361 DDIQFTMGSSPTDITYTVYRDGTIKIKEGLTFTTTEEDGVATGNHEYCEVKYTAGVSP 420
 Db 1310 DDIQFTMGSSPTDITYTVYRDGTIKIKEGLTFTTTEEDGVATGNHEYCEVKYTAGVSP 1369
 QY 421 KECVNVTPDPVQNFVQNLTGSAVGOKVTLKWDAPN 456
 Db 1370 KECVNVTPDPVQNFVQNLTGSAVGOKVTLKWDAPN 1405

RESULT 7

AA96021
 ID AAR96021 standard; protein; 450 AA.

AC AAR96021;

XX 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

DE P. gingivalis haemagglutinin hAgA Harepl product.

XX Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harepl.

XX Porphyromonas gingivalis; strain 381.

XX WO9617936-A2.

XX 13-JUN-1996.

XX 11-DEC-1995; 95WO-US016108.

XX 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX WPI; 1996-287181/29.

DR N-PSDB; AAT30645.

XX Porphyromonas gingivalis genes and proteins - used in the detection and vaccination against periodontal disease.

XX Claim 4; Page 103-104; 153pp; English.

XX HAREPL (AAR96021) is the product of the Harepl repeat unit (AAT30645) of the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of haemagglutinin hAgA (see also AAR96030). Harepl and other hAgA repeat unit products (see also AAR96022-24) can be obtd. from transformed host cells and used as vaccines to protect humans or animals against periodontal disease. Expression in Salmonella cells allows prodn. of live vaccine. Harepl-4 can also be used to detect the presence of anti-P. gingivalis antibodies and to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 450 AA;

Query Match

Best Local Similarity 98.2%; Score 2436; DB 2; Length 450;

Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 PNPNGPTTLLSBSEFNGIPASWKTIIDADGGQWCLSSGQWLTARHGTN 66
 Db 1 PNPNGPTTLLSBSEFNGIPASWKTIIDADGGQWCLSSGQWLTARHGTN 60
 QY 67 NFEQPONPNYLVTPELSIPNGGTLTFWYCAQDANYASEHYAVYASSTGNDASNFANALL 126
 Db 61 NFEQPONPNYLVTPELSIPNGGTLTFWYCAQDANYASEHYAVYASSTGNDASNFANALL 120
 QY 127 EEVLTAKTAVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHYNCSDLYNILLDDVEIK 186
 Db 121 EEVLTAKTAVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHYNCSDLYNILLDDVEIK 180
 QY 187 ANGKRAADFTETTESSTHGEAPAEWTTTIDADGGQWCLSSGQWLTARHGTN 246
 Db 181 ANGKRAADFTETTESSTHGEAPAEWTTTIDADGGQWCLSSGQWLTARHGTN 240
 QY 247 WNGMALPNPNYLSKDVGTGKVKYVAVNDGPGDHYAVMISKTGNAGDFTV 306
 Db 241 WNGMALPNPNYLSKDVGTGKVKYVAVNDGPGDHYAVMISKTGNAGDFTV 300
 QY 307 NGINKGARFGLSTEADGAKPQSVWERTVDLPAGTKYVAFRHYNCSDLYNILLDDIQT 366
 Db 301 NGINKGARFGLSTEADGAKPQSVWERTVDLPAGTKYVAFRHYNCSDLYNILLDDIQT 360
 QY 367 MGSPTPTDITYTVYRDGTIKIKEGLTFTTTEEDGVATGNHEYCEVKYTAGVSPKECVNV 426
 Db 361 MGSPTPTDITYTVYRDGTIKIKEGLTFTTTEEDGVATGNHEYCEVKYTAGVSPKECVNV 420
 QY 427 TVDPVQNFVQNLTGSAVGOKVTLKWDAPN 456
 Db 421 TVDPVQNFVQNLTGSAVGOKVTLKWDAPN 450

RESULT 8

AAW69489

ID AAW69489 standard; protein; 450 AA.

XX AAW69489;

XX 22-DEC-1998 (first entry)

DE Haemagglutinin protein hAgA, Harepl.

XX Haemagglutinin protein; periodontal disease; vaccine; hAgA.

XX Porphyromonas gingivalis.

XX US824791-A.

XX 20-OCT-1998.

XX 11-DEC-1995; 95US-00570311.

XX 08-SEP-1988; 88US-00241540.

XX 25-JAN-1991; 91US-00647119.

XX 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;

XX WPI; 1998-582627/49.

XX N-PSDB; AAV58876.

XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or protease poly:peptide(s)).

XX Claim 1; Col 121-126; 101pp; English.

XX This sequence is encoded by a Porphyromonas gingivalis gene of the invention. This sequence represents the hAgA haemagglutinin protein. The

CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease

XX SQ Sequence 450 AA;
 Query Match 98.2%; Score 2436; DB 2; Length 450;
 Best Local Similarity 99.8%; Pred. No. 2.6e-187;
 Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 7 PNPNGPTTILSEFENGIPASWKIIDADGDNNTTTPPGGTSFAGHNSAICAASSASYI 66
 DB 1 PNPNGPTTILSEFENGIPASWKIIDADGDNNTTTPPGGTSFAGHNSAICAASSASYI 60
 QY 67 NFEQPQNDVLTPELSPNGGTLTFWVCAQDANYASEHVAVYASSTGNDASFNALL 126
 DB 61 NFEQPQNDVLTPELSPNGGTLTFWVCAQDANYASEHVAVYASSTGNDASFNALL 120
 QY 127 BEVLTAKTVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNLDDVEIK 186
 DB 121 BEVLTAKTVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNLDDVEIK 180
 QY 187 ANGKRAFTTFESSTHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGCTNVASFS 246
 DB 181 ANGKRAFTTFESSTHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGCTNVASFS 240
 QY 247 WNGMALPNDNLI SKDVTGATKVKYYAVNDGFGPDHYAVMISKTGYNAGDFTVVFETP 306
 DB 241 WNGMALPNDNLI SKDVTGATKVKYYAVNDGFGPDHYAVMISKTGYNAGDFTVVFETP 300
 QY 307 NGKNGGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHNCSDLNVLDDIQT 366
 DB 301 NGKNGGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHNCSDLNVLDDIQT 360
 QY 367 MGGSPTPTDVTYVYRDGTKIKEGLTETTFEEDGVATGNHHCYCVVEKYTAGVSPKCVNV 426
 DB 361 MGGSPTPTDVTYVYRDGTKIKEGLTETTFEEDGVATGNHHCYCVVEKYTAGVSPKCVNV 420
 QY 427 TVDPVQNPVQNLGSAVGQKVTLLKWDAPN 456
 DB 421 TVDPVQNPVQNLGSAVGQKVTLLKWDAPN 450

RESULT 9
 AAR96024
 ID AAR96024 standard; protein; 439 AA.
 AC AAR96024;
 XX
 XX
 XX 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 XX
 DE P. gingivalis haemagglutinin hAgA Harep4 product.
 XX
 XX Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harep4.
 XX
 XX Porphyromonas gingivalis; strain 381.
 OS
 XX
 XX WO9617936-A2.
 FN
 XX
 XX 13-JUN-1996.
 PD
 XX
 XX 11-DEC-1995; 95WO-US016108.
 PF
 XX
 XX 09-DEC-1994; 94US-00353485.
 PR
 XX
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 EA
 XX
 XX Proguliske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 FI
 XX
 XX WPI; 1996-287181/29.
 DR

DR N-PSDB; AAT30648.
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 4; Page 114-115; 153pp; English.
 XX
 CC Harep4 (AAR96024) is the product of the Harep4 repeat unit (AAT30648) of
 CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
 CC haemagglutinin hAgA (see also AAR96030). Harep4 and other hAgA repeat
 CC unit products (see also AAR96021-23) can be obtained from transformed host
 CC cells and used as vaccines to protect humans or animals against
 CC periodontal disease. Expression in Salmonella cells allows production of live
 CC vaccine. Harep1-4 can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 XX SQ Sequence 439 AA;
 Query Match 94.8%; Score 2351; DB 2; Length 439;
 Best Local Similarity 98.2%; Pred. No. 1.7e-180;
 Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 1 GTPNPNPNGPTTILSEFENGIPASWKIIDADGDNNTTTPPGGTSFAGHNSAICA 60
 DB 1 GTPNPNPNGPTTILSEFENGIPASWKIIDADGDNNTTTPPGGTSFAGHNSAICA 60
 QY 61 SSASYNFEGPQNDVLTPELSPNGGTLTFWVCAQDANYASEHVAVYASSTGNDASN 120
 DB 61 SSASYNFEGPQNDVLTPELSPNGGTLTFWVCAQDANYASEHVAVYASSTGNDASN 120
 QY 121 FANALLEVLTAKTVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
 DB 121 FANALLEVLTAKTVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
 QY 181 DVEIHKANGKRAFTTFESSTHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGCTN 240
 DB 181 DVEIHKANGKRAFTTFESSTHGEAPAEWTTIDADGGGWLCLSSGQLGWLTAHGCTN 240
 QY 241 VVASFSWNGMALPNDNLI SKDVTGATKVKYYAVNDGFGPDHYAVMISKTGYNAGDFTV 300
 DB 241 VVASFSWNGMALPNDNLI SKDVTGATKVKYYAVNDGFGPDHYAVMISKTGYNAGDFTV 300
 QY 301 VFETPNKNGGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHNCSDLNVL 360
 DB 301 VFETPNKNGGARFGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHNCSDLNVL 360
 QY 361 DDIOFTMGGSPTPTDVTYVYRDGTKIKEGLTETTFEEDGVATGNHHCYCVVEKYTAGVSP 420
 DB 361 DDIOFTMGGSPTPTDVTYVYRDGTKIKEGLTETTFEEDGVATGNHHCYCVVEKYTAGVSP 420
 QY 421 KECNVTVTVDVQNPVQNL 439
 DB 421 KVCNVTVTINPTQNPVQNL 439
 RESULT 10
 AAW69492
 ID AAW69492 standard; protein; 439 AA.
 XX
 XX AAW69492;
 AC
 XX
 XX 22-DEC-1998 (first entry)
 DT
 XX
 XX Haemagglutinin protein hAgA, Harep4.
 DE
 XX
 XX Haemagglutinin protein; periodontal disease; vaccine; hAgA.
 KW
 XX
 XX Porphyromonas gingivalis.
 OS
 XX
 XX US5824791-A.
 FN
 XX
 XX 20-OCT-1998.
 PD

```

XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX
XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulskie-Fox A, Lepine G;
XX WPI; 1998-582627/49.
XX N-PSDB; AAV58879.
XX
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly.peptide(s)).
XX
XX Claim 1; Col 139-144; 101pp; English.
XX
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX invention. This sequence represents the hgaA haemagglutinin protein. The
XX polypeptides are used to produce antibodies to organisms associated with
XX periodontal disease. The antibodies are also used in purification and
XX identification procedures. The genes and polypeptides are used as
XX vaccines against periodontal disease
XX
XX Sequence 439 AA;
XX
Query Match 94.8%; Score 2351; DB 2; Length 439;
Best Local Similarity 98.2%; Pred. No. 1.7e-180;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 GTPNPNPNPGTTLSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICA 60
Db 1 GTPNPNPNPGTTLSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICV 60
QY 61 SSASYINFGQPNDNYLTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 61 SSASYINFGQPNDNYLTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEEVLTAKTVTAPEARTGTVOGTWYKTVQVLPAGTKYVAFRHFQCTDFWNL 180
Db 121 FANALLEEVLTAKTVTAPEARTGTVOGTWYKTVQVLPAGTKYVAFRHFQCTDFWNL 180
QY 181 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLWLTAGHGTN 240
Db 181 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLWLTAGHGTN 240
QY 241 VVASFSWNGMALNPDNYLISKDVTGATKVKYIYVNDGPPGDHYAVMISKTGTNAGDFTV 300
Db 241 VVASFSWNGMALNPDNYLISKDVTGATKVKYIYVNDGPPGDHYAVMISKTGTNAGDFTV 300
QY 301 VVEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
Db 301 VVEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
QY 361 DDIOFTMGSPPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
Db 361 DDIOFTMGSPPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
QY 421 KVCNVNTPDVPQNPQN 439
Db 421 KVCNVNTPDVPQNPQN 439
RESULT 11
AAR96033
ID AAR96033 standard; protein; 1687 AA.
XX
XX AAR96033;
XX
XX 16-OCT-2003 (revised)
DT

```

```

DT 04-SEP-1996 (first entry)
XX
XX P. gingivalis haemagglutinin hage.
XX
XX Haemagglutinin; hage; periodontal disease; vaccine; antibody.
XX
XX Porphyromonas gingivalis; strain FDC381.
XX
XX WO9617936-A2.
XX
XX 13-JUN-1996.
XX
XX 11-DEC-1995; 95WO-USO16108.
XX
XX 09-DEC-1994; 94US-00353485.
XX
XX (UYFL ) UNIV FLORIDA.
XX
XX (UABR-) UAB RES FOUND.
XX
XX Progulskie-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX WPI; 1996-287181/29.
XX N-PSDB; AAT30456.
XX
XX Porphyromonas gingivalis genes and proteins - used in the detection and
XX vaccination against periodontal disease.
XX
XX Claim 5; Page 138-143; 153pp; English.
XX
XX P. gingivalis 381 haemagglutinin hage (AAR96033) was identified as the
XX product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
XX The haemagglutinin can be obtd. from transformed host cells and used as a
XX vaccine to protect humans or animals against periodontal disease.
XX Expression in Salmonella cells allows prodn. of a live vaccine. The
XX haemagglutinin can also be used to detect the presence of anti-P.
XX gingivalis antibodies and to raise monoclonal antibodies for diagnostic
XX appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 1687 AA;
XX
Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 5.2e-158;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;
QY 3 PNPENPNPGTTLSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICASS 62
Db 935 PNPENPNPGTTLSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICV 991
QY 63 ASY-INFGQPNDNYLTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 121
Db 992 ESFGLGGIGVLPDNYLITPALDLPLNGGKLTWVCAQDANYASEHYAVYASSTGNDASN 1051
QY 122 ANALLEEVLTAKTVTAPEARTGTVOGTWYKTVQVLPAGTKYVAFRHFQCTDFWNL 181
Db 1052 TNALLEETITAGK-VRSPEAING-RIQGTWROKTVLDPAGTKYVAFRHFQCTDFWNL 1109
QY 182 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLWLTAGHGTN 241
Db 1110 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQGLWLTAGHGTN 1169
QY 242 VASFSWNGMALNPDNYLISKDVTGATKVKYIYVNDGPPGDHYAVMISKTGTNAGDFTV 301
Db 1170 VASFSWNGMALNPDNYLISKDVTGATKVKYIYVNDGPPGDHYAVMISKTGTNAGDFTV 1229
QY 302 FEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 361
Db 1230 FEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 1289
QY 362 DIQFTMGSPPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 421
Db 1290 DIQFTMGSPPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVVKYTAGVSP 1349
QY 422 ECVNVTVDVPQNPQNLTGSAVGKVTWKDAPN 456

```

```

|||||:::|||||:::|||||:::|||||:::|||||:::
Db 1350 ECVNVTINPTQFNPKLKAQPDGGDVVLKWEAPS 1384

RESULT 12
AAW69495
ID AAW69495 standard; protein; 1687 AA.
XX AC AAW69495;
XX XX
DT 22-DEC-1998 (first entry)
XX DE Haemagglutinin protein hage.
XX DE Haemagglutinin protein; periodontal disease; vaccine; hage.
XX KW Porphyromonas gingivalis.
XX OS Porphyromonas gingivalis.
XX PN US5824791-A.
XX PD 20-OCT-1998.
XX PF 11-DEC-1995; 95US-00570311.
XX PR 08-SEP-1988; 88US-00241640.
XX PR 25-JAN-1991; 91US-00647119.
XX PR 09-DEC-1994; 94US-00353485.
XX PA (UYFL) UNIV FLORIDA.
XX PA (UABR-) UAB RES FOUND.
XX PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulsk-Fox A, Lepine G;
XX DR WPI; 1998-582627/49.
XX DR N-PSDB; AAV58881.
XX FT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX FT protease poly:peptide(s).
XX PS Claim 1; Col 167-182; 101pp; English.
XX CC This sequence is encoded by a Porphyromonas gingivalis gene of the
XX CC invention. This sequence represents the hage haemagglutinin protein. The
XX CC polypeptides are used to produce antibodies to organisms associated with
XX CC periodontal disease. The antibodies are also used in purification and
XX CC identification procedures. The genes and polypeptides are used as
XX CC vaccines against periodontal disease
XX SQ Sequence 1687 AA;

Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 5.2e-158;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTISEFENGIPASWKTIADGQGNWTTTPPGGTSFAGHNSAICASS 62
Db 935 PNPENPNPGTTTISEFENGIPASWKTIADGQGNWTTTPPGGTSFAGHNSAICASS 991
QY 63 ASY- INFEGPQPNYLVTPELSLPNGGTLTFWCAQDANYASHEVAYASSTGNDASNF 121
Db 992 ESFGLGGIGVLTFPNYLITPALDIPNGGKLTFWCAQDANYASHEVAYASSTGNDASNF 1051
QY 122 ANALLEEVLTAKTVVTAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRFGCTDFFWINLD 181
Db 1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRFGQSTDFYIDL 1109
QY 182 DVEIKANGKRADFTETTESSTHGEAPAEWTIIDADGGQGWHLCLSSGQLGHLTAHGNTV 241
Db 1110 EVEIKANGKRADFTETTESSTHGEAPAEWTIIDADGGQGWHLCLSSGQLGHLTAHGNTV 1169
QY 242 VASFSWNGMALNPNYLISKDVATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDTTV 301
Db 1170 VASFSWNGMALNPNYLISKDVATKVKYIYAVNDGFGPDHYAVMISKTGTNAGDTTV 1229

|||||:::|||||:::|||||:::|||||:::|||||:::
Db 1350 ECVNVTINPTQFNPKLKAQPDGGDVVLKWEAPS 1384

RESULT 13
AAR70188
ID AAR70188 standard; protein; 1704 AA.
XX AC AAR70188;
XX XX
DT 25-MAR-2003 (revised)
DT 21-SEP-1995 (first entry)
XX DE Arg-gingipain-2 prepolyprotein.
XX DE Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
XX KW arginine-specific protease.
XX OS Porphyromonas gingivalis.
XX XX
XX FH Key Location/Qualifiers
XX FT Protein 228..719
XX FT /label= Protease
XX FT /note= "corresponds to Arg-gingipain-1"
XX FT Region 720..1091
XX FT /label= Hemagglutinin
XX FT Region 1092..1429
XX FT /label= Hemagglutinin
XX FT Region 1430..1704
XX FT /label= Hemagglutinin
XX PN W09507286-A1.
XX XX
XX PD 16-MAR-1995.
XX PF 09-SEP-1994; 94WO-USO10283.
XX PR 10-SEP-1993; 93US-00119361.
XX PR 21-OCT-1993; 93US-00141324.
XX PR 24-JUN-1994; 94US-00265441.
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX XX
XX FI Travis J, Potempa J, Barr PJ, Pavloff N;
XX XX
XX DR WPI; 1995-123373/16.
XX DR N-PSDB; AAQ83489.
XX XX
XX PT DNA encoding Arg-gingipain proteins - used to develop prods. for
XX FT detection, treatment and prevention of periodontal disease.
XX XX
XX PS Disclosure; Page 70-77; 89pp; English.
XX CC
XX CC A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
XX CC 2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
XX CC (ATCC 53373). The sequences of the proteins were used to design PCR
XX CC primers and probes to isolate AG DNA. Lambda DASH and Lambda ZAP
XX CC libraries were screened with a probe based on amino acids 11-22 of the AG
XX CC protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
XX CC is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
XX CC correct PN field.)
XX SQ Sequence 1704 AA;

```

| Query Match | 34.0% | Score 2082; | DB 2; | Length 1704; |
|-----------------------|---|---|-----------|--------------|
| Best Local Similarity | 86.2% | Fred. No. 5.2e-158; | | |
| Matches 392; | Conservative 17; | Mismatches 40; | Indels 6; | Gaps 4; |
| QY | 3 | PNPNPNPQTTILTSRPFENGIPASWKTIIDADGDGNNWTTTPPPGGTSFAGHNSAICASS | 62 | |
| Db | 952 | PNPNPNPQTTILTSRPFENGIPASWKTIIDADGDGHCWKFGNAPG---IAGYNSNGCVYS | 1008 | |
| QY | 63 | ASY-INFEQGPNPNYLVTPELSPLNPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF | 121 | |
| Db | 1009 | ESFGLGGIGVLTDPNYLITPALDILPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF | 1068 | |
| QY | 122 | ANALLEVLTAKTVTAPEAIRGTRVOGTWVKTVQLPAGTKYVAFRHFGCTDFFWNLND | 181 | |
| Db | 1069 | TNALLETITAKG-VRSPEAIRG-RIOGTWRQKTVDLUPAGTKYVAFRHFGCTDFFWNLND | 1126 | |
| QY | 182 | DVEIKANGKRADETTFESSTHGEAPAEWTTIIDADGDGQGWLCILSSQQLGWLTAHGCTNV | 241 | |
| Db | 1127 | DVEIKANGKRADETTFESSTHGEAPAEWTTIIDADGDGQGWLCILSSQQLDWLTAHGCTNV | 1186 | |
| QY | 242 | VASFWSNMGALNPNYLISKDVTGATKVKYKYAVVNDGFGPDGHYAVMLSKTGTNAGDFTVV | 301 | |
| Db | 1187 | VASFWSNMGALNPNYLISKDVTGATKVKYKYAVVNDGFGPDGHYAVMLSKTGTNAGDFTVV | 1246 | |
| QY | 302 | FEETPNGINKGGARFGLSTADGAKPOSVMIERVTDLUPAGTKYVAFRHYNCSDLNYILLD | 361 | |
| Db | 1247 | FEETPNGINKGGARFGLSTANGAKPOSVMIERVTDLUPAGTKYVAFRHYNCSDLNYILLD | 1306 | |
| QY | 362 | DIQFTMGGSPTDYYTVVTRDGTIKIEGLTETTTFEEDGVTATGNHEYCVVEKYYTAGVSPK | 421 | |
| Db | 1307 | DIQFTMGGSPTDYYTVVTRDGTIKIEGLTETTTFEEDGVTATGNHEYCVVEKYYTAGVSPK | 1366 | |
| QY | 422 | ECNVNVTDPQVQPNVQNLGTSVAGQKVTLLKWDAPN | 456 | |
| Db | 1367 | ECNVNVTINQFNPVKNLKAQPDGDDVLLKWEAPS | 1401 | |
| RESULT 14 | | | | |
| AAW34843 | | | | |
| ID | AAW34843 standard; protein; 1704 AA. | | | |
| AC | AAW34843; | | | |
| DT | 03-JUN-1998 (first entry) | | | |
| XX | Arg-gingipain high molecular weight prepolyprotein sequence. | | | |
| XX | Arg-specific gingipain protease; gingivitis; periodontal disease; | | | |
| KW | vaccine; infection. | | | |
| OS | Porphyromonas gingivalis. | | | |
| XX | Key Location/Qualifiers | | | |
| FT | 1. .227 | | | |
| FT | /note= "precursor protein" | | | |
| XX | W09734629-A1. | | | |
| PN | XX | | | |
| PD | 25-SEP-1997. | | | |
| XX | 21-MAR-1997; 97WO-US004635. | | | |
| PF | 22-MAR-1996; 96US-0013945P. | | | |
| XX | (UYGE-) UNIV GEORGIA RES FOUND INC. | | | |
| XX | (MORE-) MOREHOUSE SCHOOL MEDICINE. | | | |
| PI | Potempa J, Travis J, Genco C; | | | |
| XX | WFI; 1997-479993/44. | | | |
| DR | N-PSDB; AAT93872. | | | |
| XX | | | | |

Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
useful for protecting animals and humans from gingivitis and periodontal
diseases.

Disclosure; Page 68-73; 95pb; English.

The present sequence represents an arginine-specific protease of
Porphyromonas gingivalis. The following peptides, derived from Arg- and
lys-specific high molecular weight proteases, offer protection against
infection: YTVYVRDGG KEGGLTATTE DGVATGNHE YCVEKTAGS VSPKVC (I);
YTPVEEKGQ RMIVAVAKY (II); QLPFIPEFVAC VNGDELFEMP CFAELMRAQ (III);
GEPNYPQVS NLATVIOQOK VILMKADPSTK (IV); GNHEYCEVK YTAGVSPKC KDVTV (V);
RMWNVBEGR YTPVEEKGQ (VI); TFAGEDTYPK RMWNVBEGR (VII); YTVYVRDGT KKEGLTATT
TKIKEGLTATTFEEDGVATG NMEYCVCCVY TAGUSPKVC (VIII); YTVYVRDGT KKEGLTATT
BEDG (IX); RDGKIKEGL TATTFEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
& KWDAPNGTN PNP PNPNPN PGTTLUSE (XII); and YTPVEEKGQ RMIVAVAKY
(XIII). They are used in vaccines to protect animals, including humans,
from gingivitis and/or periodontal diseases

Sequence 1704 AA;

Query Match 84.0%; Score 2082; DB 2; Length 1704;
Best Local Similarity 86.2%; Pred.No.5.2e-158;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTITLSESPENGIPASWKTIADGDGNNWTTPPPGGTSFAGHNSAICASS 62
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
952 PNPENPNPGTITLSESPENGIPASWKTIADGDGHGWKPNAG---IAGVNSNGCVYS 1008
QY 63 ASY-INFEQGPNPDNYLTPELSLPLNGGTLTPFWCAQDANYASEHYAVYASSTGNDASNF 121
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1009 ESFGLGIGVLPDNYLIPTALDLPNGGKLTPWCAQDANYASEHYAVYASSTGNDASNF 1068
QY 122 ANALLEEVLTAKTVTAPEAIRGTVGCTGYOXTVOLPAGTKYAFRHFGCTDFEWINLD 181
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1069 THALLEETITAKG-VRSPPAIRG-RIQGTWRQKTVDLPAGTKYAFRHFPQSTMDFIIDLD 1126
QY 182 DVEIKANGKRADFTTFESSTHGEAPAERTTIDADGDGGWLCLSSGQLGMLTAHGGINV 241
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1127 EVEIKANGKRADFTTFESSTHGEAPAERTTIDADGDGGWLCLSSGQLDLWLTAGGINV 1186
QY 242 VASFWSWGVALPDNYLVISKVTGATKYKYVAVNDGPDPDHAYVMISKTGYNAGDFTVV 301
Db 1187 VASFWSWGVALPDNYLVISKVTGATKYKYVAVNDGPDPDHAYVMISKTGYNAGDFTVV 1246
QY 302 FEETPNGINKGARFGLSTEADGAKPSQSWIERTVDLPAGTKYAFRVHNCSDLNLYILD 361
Db 1247 FEETPNGINKGARFGLSTEANGAKPSQSWIERTVDLPAGTKYAFRVHNCSDLNLYILD 1306
QY 362 DIQFTMGGSPTDTYTYYVRDGTKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSPK 421
Db 1307 DIQFTMGGSPTDTYTYYVRDGTKIKEGLTTFEEDGVATGNHEYCVVKYTAGVSPK 1366
QY 422 ECNVNTVDVPQNVPQNLTGSAVGQKVLTKWDAPN 456
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1367 ECVNVTINFTQNPVNKLKAQPDGDDVLKWEAPS 1401

RESULT 15
AAy67396
ID AAY67396 standard; protein; 1704 AA.
XX
XX AAY67396;
XX
XX
XX 25-APR-2000 (first entry)
XX
XX Arg-gingipain-2 amino acid sequence.
XX
XX Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
XX immunogenic component; vaccine; inflammatory response; tissue damage;
XX periodontal disease.
XX
XX Porphyromonas gingivalis.

XX Key Location/Qualifiers
FH Region 229..719
FT /note= "Amino acids 229-719 are specifically claimed"
FT Region 720..1185
FT /note= "Amino acids 720-1185 are specifically claimed"
XX
PN US6017532-A.
XX
XX 25-JAN-2000.
XX
XX 08-NOV-1994; 94US-00336308.
XX
XX 10-SEP-1993; 93US-00119361.
XX 24-JUN-1994; 94US-00265441.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX Potempa JS, Travis J;
XX
XX MPI; 2000-136659/12.
XX
XX N-PSDB; AAZ60181.
XX
XX New Porphyromonas gingivalis arginine-specific protease preparation
PT useful for preparing vaccines against periodontal disease and for
PT screening for Arg-gingipain inhibitors.
XX
XX
XX Claim 1; Col 29-42; 55pp; English.
XX
XX This sequence represents a Porphyromonas gingivalis arginine-specific
CC proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
CC Gingipain-2 consists of a 50kD protease component non-covalently
CC associated with a 44kD haemagglutinin component. The proteinase is
CC stimulated by glycine containing peptides and glycine analogues. It is
CC inhibited by cysteine protease group specific inhibitors. The protease
CC preparation can be used in immunogenic compositions and vaccines against
CC inflammatory response and tissue damage caused by P. gingivalis in
CC periodontal disease. It can also be used to screen for agents that
CC modulate Arg-gingipain proteinase activity inhibitors
XX
XX Sequence 1704 AA;
SQ

Query Match 84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 5.2e-158;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;
QY 3 PNPENPNPCTTTLSSEFENGIPASWKTIDADGDNNTTTTPPGGTSPAGHNSAICASS 62
DB PNPENPNPCTTTLSSEFENGIPASWKTIDADGDNNTTTTPPGGTSPAGHNSAICASS 1008
QY 63 ASY-INFEQPNPNLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
DB ESFGLGGIGVLTPTDNYLITPALDPLNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068
QY 122 ANALLEEVLTAKVTVAPEAIRTRVGSTWYQVQLPAGTKYVAFRHFCTDFFWNLID 181
DB TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDLD 1126
QY 182 DVEIKANGKRADETTFESSTHGEAPAEWTTIDADGGQGWLCISGQLGWLTAHGNTV 241
DB EVIIKANGKRADETTFESSTHGEAPAEWTTIDADGGQGWLCISGQLGWLTAHGNTV 1186
QY 242 VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKGTGNTAGDFTVV 301
DB VASFSWNGMALPNPNYLISKDVTGATKVKYKYAVVNDGFGPDHYAVMISKGTGNTAGDFTVV 1246
QY 302 FEETPGINKGARFGLSTADGAKPOSWIERTVLDPAGTKYVAFRHYNCSDINYLID 361
DB FEETPGINKGARFGLSTADGAKPOSWIERTVLDPAGTKYVAFRHYNCSDINYLID 1306
QY 362 DIQFTMGGSPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHGYCVKVTAGVSPK 421
DB DIQFTMGGSPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHGYCVKVTAGVSPK 1366

QY 422 ECVNVTVDVQFNPVQNLTGSAVGQKVTLKWDAPN 456
DB 1367 ECVNVTINVTQFNPVKLKAQPDGGDVVLKWEAPS 1401

Search completed: May 18, 2004, 11:42:43
Job time : 48.0287 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:37:00 ; Search time 11.906 seconds
(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-20
Perfect score: 2480
Sequence: 1 GTPNPNPNPFGTTLSSEF.....QNLGSAVGQVKTLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 2480 | 100.0 | 2628 | 2 T28651 | hemagglutinin A - |
| 2 | 2082 | 84.0 | 1704 | 2 A55426 | gingipain R (EC 3. |
| 3 | 2059 | 83.0 | 1526 | 2 S45763 | gingipain R (EC 3. |
| 4 | 2034.5 | 82.0 | 1732 | 2 T30835 | lysine-specific cy |
| 5 | 266 | 10.7 | 991 | 2 I40229 | arginyl endopeptid |
| 6 | 139.5 | 5.6 | 1052 | 2 AF2959 | conserved hypothet |
| 7 | 139.5 | 5.6 | 1341 | 2 H98323 | hypothetical prote |
| 8 | 132.5 | 5.3 | 691 | 2 B75622 | hypothetical prote |
| 9 | 131.5 | 5.3 | 2468 | 2 A83412 | hypothetical prote |
| 10 | 128.5 | 5.2 | 1248 | 2 C98974 | autolysin [importe |
| 11 | 128 | 5.2 | 4936 | 2 A2515 | hypothetical prote |
| 12 | 127.5 | 5.1 | 1684 | 2 S10789 | amylase A-180 - al |
| 13 | 127 | 5.1 | 713 | 2 B75489 | hypothetical prote |
| 14 | 126 | 5.1 | 1441 | 2 A85685 | prophage p11 prote |
| 15 | 126 | 5.1 | 1649 | 2 C86822 | hypothetical prote |
| 16 | 124.5 | 5.0 | 1904 | 2 T13256 | tail-host specific |
| 17 | 124.5 | 5.0 | 5291 | 2 F90696 | hypothetical prote |
| 18 | 123.5 | 5.0 | 5188 | 2 B85547 | probable RTX fami |
| 19 | 122.5 | 4.9 | 1274 | 2 T10729 | transferrin-like p |
| 20 | 121.5 | 4.9 | 465 | 2 A47023 | S-layer protein - |
| 21 | 121.5 | 4.9 | 3624 | 2 A08035 | large repetitive p |
| 22 | 121 | 4.9 | 1090 | 2 S59077 | cellulose 1,4 beta |
| 23 | 120 | 4.8 | 908 | 2 AE2254 | hypothetical prote |
| 24 | 120 | 4.8 | 2783 | 2 T34416 | hypothetical prote |
| 25 | 119 | 4.8 | 715 | 2 JC4908 | alkaline serine pr |
| 26 | 119 | 4.8 | 1873 | 2 T30944 | surface protein pr |
| 27 | 118.5 | 4.8 | 926 | 2 D86897 | hypothetical prote |
| 28 | 118.5 | 4.8 | 1385 | 2 T18213 | parasporal crystal |
| 29 | 118.5 | 4.8 | 2817 | 2 B97033 | uncharacterized pr |

| | | | | | |
|----|-------|-----|------|----------|--------------------|
| 30 | 118.5 | 4.8 | 4199 | 2 S76412 | hypothetical prote |
| 31 | 118 | 4.8 | 1461 | 2 E90696 | hypothetical prote |
| 32 | 118 | 4.8 | 1461 | 2 A85547 | hypothetical prote |
| 33 | 117.5 | 4.7 | 1034 | 2 T30551 | beta-galactosidase |
| 34 | 116.5 | 4.7 | 1635 | 2 A10452 | hemolysin [importe |
| 35 | 115.5 | 4.7 | 702 | 2 S48753 | major surface prot |
| 36 | 115.5 | 4.7 | 875 | 2 AF0472 | probable outer mem |
| 37 | 115.5 | 4.7 | 1118 | 1 A49724 | protein-tyrosine-p |
| 38 | 115.5 | 4.7 | 1282 | 2 JC4393 | microbial collagen |
| 39 | 115.5 | 4.7 | 1939 | 2 D97316 | probable S-layer p |
| 40 | 115 | 4.6 | 821 | 2 AD1507 | probable secreted |
| 41 | 115 | 4.6 | 1651 | 2 JC1340 | outer membrane pro |
| 42 | 115 | 4.6 | 3029 | 2 S76109 | hypothetical prote |
| 43 | 114 | 4.6 | 872 | 2 S49541 | cellulase - Cellul |
| 44 | 113 | 4.6 | 1345 | 2 H90975 | hypothetical prote |
| 45 | 112.5 | 4.5 | 938 | 2 AF1772 | internalin-like pr |

ALIGNMENTS

RESULT 1

T28651
hemagglutinin A - Porphyromonas gingivalis
C:Species: Porphyromonas gingivalis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001
C:Accession: T28651
R:Han, N.; Whitlock, J.; Progulskie-Fox, A.
Infect. Immun. 64, 4000-4007, 1996
A:Title: The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381 contains four
A:Reference number: Z20494; MUID: 97047672; PMID: 8926061
A:Accession: T28651
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2628 <HAN>
A:Cross-references: EMBL:U41807; NID:gl552410; PID:gl1469916; PIDN:AAB17128.1
C:Genetics:
A:Gene: haga

| | | | | | | | |
|-----------------------|--------|--|-----------|------------|----|--------|-------|
| Query Match | 100.0% | Score | 2480; | DB | 2; | Length | 2628; |
| Best Local Similarity | 100.0% | Pred. No. | 1.4e-162; | Mismatches | 0; | Indels | 0; |
| Matches | 456; | Conservative | 0; | Indels | 0; | Gaps | 0; |
| QY | 1 | GTPNPNPNPFGTTLSSEFENGIPASWKITDADGDGNNWTTTPPPGGTSFAGHNSAICA | 60 | | | | |
| DB | 950 | GTPNPNPNPFGTTLSSEFENGIPASWKITDADGDGNNWTTTPPPGGTSFAGHNSAICA | 1009 | | | | |
| QY | 61 | SSASVINFEQNPDPNYLVTPELSLPNGGTLTFWVCAQADANYASHYAVYASSTGNDASN | 120 | | | | |
| DB | 1010 | SSASVINFEQNPDPNYLVTPELSLPNGGTLTFWVCAQADANYASHYAVYASSTGNDASN | 1069 | | | | |
| QY | 121 | FANALLEVLTAKTVVTAPPAIRGTRVOGTWYQKTVQLPAGTKYVAFRHFCTDFFWNL | 180 | | | | |
| DB | 1070 | FANALLEVLTAKTVVTAPPAIRGTRVOGTWYQKTVQLPAGTKYVAFRHFCTDFFWNL | 1129 | | | | |
| QY | 181 | DDVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQGWMLCSSQQLGWLTAHGGTN | 240 | | | | |
| DB | 1130 | DDVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQGWMLCSSQQLGWLTAHGGTN | 1189 | | | | |
| QY | 241 | VVASFSWNGMALNPDNYLISKDVTGATKVKYVAVNDGFGPDGHVAVMLSKGTGNAGDTV | 300 | | | | |
| DB | 1190 | VVASFSWNGMALNPDNYLISKDVTGATKVKYVAVNDGFGPDGHVAVMLSKGTGNAGDTV | 1249 | | | | |
| QY | 301 | VFETPNKINGGARFGLSTEADGAKPQSVIERTVLPAGTKYVAFRHYNCSDLNYILL | 360 | | | | |
| DB | 1250 | VFETPNKINGGARFGLSTEADGAKPQSVIERTVLPAGTKYVAFRHYNCSDLNYILL | 1309 | | | | |
| QY | 361 | DDIQTWGGSPPTDYTYTVYVRDGTKEGTTETTTFEDGVATGNHCEYCVKVTAGVSP | 420 | | | | |
| DB | 1310 | DDIQTWGGSPPTDYTYTVYVRDGTKEGTTETTTFEDGVATGNHCEYCVKVTAGVSP | 1369 | | | | |
| QY | 421 | KECVNVTVDVQFNPVQNLGTSAGVQKVTLKWDAPN | 456 | | | | |

Db 1370 KCVNVTVDVQFNPVQNLGSAVGQKVTWKWDAPN 1405

RESULT 2
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N:Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; H3P; R
C:Species: Porphyromonas gingivalis
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C:Accession: A55426; D53113
R:Payloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, E
J. Biol. Chem. 270, 10077-1010, 1995
A:Title: Molecular cloning and structural characterization of the Arg-gingipain proteinase
A:Reference number: A55426; MUID:95138080; PMID:7836351
A:Accession: A55426
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1704 <PAV>
A:Cross-references: GB:U15282; NID:9557067; PIDN:AAA69539.1; PID:9557068
R:Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A:Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A:Reference number: A53113; MUID:94103245; PMID:8276827
A:Accession: D53113
A:Status: preliminary
A:Molecule type: protein
A:Residues: 228-249 <PIK>
A:Experimental source: H66
A:Note: sequence extracted from NCBI backbone (NCBI:141694)
C:Keywords: cysteine proteinase; hydrolase

Query Match 84.0%; Score 2082; DB 2; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.2e-135;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

Qy 3 PNPFPNPGTTTILSESPENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASS 62
Db 952 PNPFPNPGTTTILSESPENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASS 62
Qy 63 ASY-INFEQPONPNYLTPNLSLTPNGGTLFWCAQDANYASHYAVYASSTGNDASNF 121
Db 1009 ESFGLGGIGVLTDPNYLITPALDLPNGGKLTFWCAQDANYASHYAVYASSTGNDASNF 1068
Qy 122 ANALLEVLTAKTVTAPALRGTRVQGTWVQKTVQVLPAGTKYVAFRHFGCTDFFWINLD 181
Db 1069 TNALEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFGCTDFFWINLD 1126
Qy 182 DVEIKANGKRAADFTETPSSHTGAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 241
Db 1127 EVEIKANGKRAADFTETPSSHTGAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 1186
Qy 242 VASFSWNGMALPNPNYLISKDVTGATKYYKYAVYVNDGPGDHYAVMISKTGTNAGDTTV 301
Db 1187 VASFSWNGMALPNPNYLISKDVTGATKYYKYAVYVNDGPGDHYAVMISKTGTNAGDTTV 1246
Qy 302 FEETPNGINKGARGFLSTADGAKPOSWIERTVLDLPAGTKYVAFRHYNCSDLNYLLD 361
Db 1247 FEETPNGINKGARGFLSTADGAKPOSWIERTVLDLPAGTKYVAFRHYNCSDLNYLLD 1306
Qy 362 DIQFTMGSPPTDITYTVYVDRGDKIKEGLTETTFEEDGVTAGNHCYVEVKYTAGVSPK 421
Db 1307 DIQFTMGSPPTDITYTVYVDRGDKIKEGLTETTFEEDGVTAGNHCYVEVKYTAGVSPK 1366
Qy 422 ECNVNTVDVQFNPVQNLGSAVGQKVTWKWDAPN 456
Db 1367 ECNVNTVDVQFNPVQNLGSAVGQKVTWKWDAPN 1401

RESULT 3
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C:Species: Porphyromonas gingivalis
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C:Accession: S49763
R:Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A:Description: Cloning, sequence analysis and expression in Escherichia coli of prpR1 of
A:Reference number: S49763
A:Accession: S49763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <ADU>
A:Cross-references: EMBL:X82680
C:Genetics:
A:Gene: prpR1
C:Keywords: cysteine proteinase; hydrolase

Query Match 83.0%; Score 2059; DB 2; Length 1526;
Best Local Similarity 85.5%; Pred. No. 7.4e-134;
Matches 389; Conservative 17; Mismatches 43; Indels 6; Gaps 4;

Qy 3 PNPFPNPGTTTILSESPENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASS 62
Db 954 PNPFPNPGTTTILSESPENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASS 62
Qy 63 ASY-INFEQPONPNYLTPNLSLTPNGGTLFWCAQDANYASHYAVYASSTGNDASNF 121
Db 1011 ESFGLGGIGVLTDPNYLITPALDLPNGGKLTFWCAQDANYASHYAVYASSTGNDASNF 1070
Qy 122 ANALLEVLTAKTVTAPALRGTRVQGTWVQKTVQVLPAGTKYVAFRHFGCTDFFWINLD 181
Db 1071 TNALEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFGCTDFFWINLD 1128
Qy 182 DVEIKANGKRAADFTETPSSHTGAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 241
Db 1129 EVEIKANGKRAADFTETPSSHTGAPAEWTTIDADGQGWCLSSGQGLWLTAGHGTNV 1188
Qy 242 VASFSWNGMALPNPNYLISKDVTGATKYYKYAVYVNDGPGDHYAVMISKTGTNAGDTTV 301
Db 1189 VASFSWNGMALPNPNYLISKDVTGATKYYKYAVYVNDGPGDHYAVMISKTGTNAGDTTV 1248
Qy 302 FEETPNGINKGARGFLSTADGAKPOSWIERTVLDLPAGTKYVAFRHYNCSDLNYLLD 361
Db 1249 FEETPNGINKGARGFLSTADGAKPOSWIERTVLDLPAGTKYVAFRHYNCSDLNYLLD 1308
Qy 362 DIQFTMGSPPTDITYTVYVDRGDKIKEGLTETTFEEDGVTAGNHCYVEVKYTAGVSPK 421
Db 1309 DIQFTMGSPPTDITYTVYVDRGDKIKEGLTETTFEEDGVTAGNHCYVEVKYTAGVSPK 1368
Qy 422 ECNVNTVDVQFNPVQNLGSAVGQKVTWKWDAPN 456
Db 1369 KCVNVTVDVQFNPVQNLGSAVGQKVTWKWDAPN 1403

RESULT 4
T30836
lysine-specific cysteine proteinase porphypain (EC 3.4.22.-) - Porphyromonas gingivalis
N:Alternate names: lysine-specific cysteine proteinase 1, 60K
C:Species: Porphyromonas gingivalis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30836; T30837; T30526; A53113
R:Barokoy-Gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Proquleske-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A:Title: Analysis of the prtp gene encoding porphypain, a cysteine proteinase of Porphy
A:Reference number: Z20895; MUID:96213011; PMID:8631659
A:Accession: T30836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1732 <BAR>
A:Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AAB06565.1
R:Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z20896
A:Accession: T30837
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A:Cross-references: EMBL:U75366; NID:G2182811; PID:G2182812; PIDN:AABG0809.1
R:Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A:Title: IS195, an insertion sequence-like element associated with protease genes in Por
phyromonas gingivalis
A:Reference number: Z20844; MUID:98298016; PMID:9632363
A:Accession: T30526
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A:Cross-references: EMBL:AF017059; NID:G2738802; PID:G2738803; PIDN:AAC26523.1
R:Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A:Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolates
A:Reference number: A53113; MUID:94103245; PMID:8276827
A:Accession: A53113
A:Status: preliminary
A:Molecule type: protein
A:Residues: 229-249 <PIK>
A:Experimental source: H66
A:Note: sequence extracted from NCBI backbone (NCBIP:141690)
C:Genetics:
A:Gene: prtP, prtK
C:Keywords: cysteine proteinase; hydrolase

Query Match 82.0%; Score 2034.5; DB 2; Length 1732;
Best Local Similarity 84.3%; Pred. No. 4.3e-132;
Matches 387; Conservative 19; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNNPCT-TTISESPENGIPASWKTIDADGNNWTTTPPGTSGAGNSAIC 59
Db 969 GTPNPNNPNNPPTLSESPENGIPASWKTIDADGNGHGWKPNAGP---TAGNSNGC 1025

QY 60 ASSASY-INFGQPNDNVLVPELSPNGGTLTFWCAQADANYASEHVYASSTGDA 118
Db 1026 VYSEFGLGGIGLVLPDNLVLPALDLPNGGLTFWCAQADANYASEHVYASSTGDA 1085

QY 119 SNFANALLEVLTAKTVTAPAIRTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWI 178
Db 1086 SNFTNALLEETITAGK-VRSKPAIRG-RIQGTWRQKTVLPAGTKYVAFRHFQSTDMFYI 1143

QY 179 NLDDVEIKANGKRAOPTTFESSSTGEAPAEWTTTIDADGGQGWCLSSGQLGWLTAHGG 238
Db 1144 DUDEVEIKANGKRAOPTTFESSSTGEAPAEWTTTIDADGGQGWCLSSGQLDMLTAHGG 1203

QY 239 TNVASFSGWNGMALPNPNVLIISKDVTGATKXYAYVNDGFGPDHYAVMISKTGTNAGDF 298
Db 1204 SNVVSFSWNGMALPNPNVLIISKDVTGATKXYAYVNDGFGPDHYAVMISKTGTNAGDF 1263

QY 299 TVVFETPNNGKNGARFGLSTEADGAKPQSVWIERTVLPAGTKYVAFRHYNCSDLNYI 358
Db 1264 TVVFETPNNGKNGARFGLSTEANGAKPQSVWIERTVLPAGTKYVAFRHYNCSDLNYI 1323

QY 359 LLDDIQTMTGSGPTDVTYTVYRGTGKIKEGLTTFTEEDGATGNHEVCVEVKYTAGV 418
Db 1324 LLDDIQTMTGSGPTDVTYTVYRGTGKIKEGLTTFTEEDGATGNHEVCVEVKYTAGV 1383

QY 419 SPKECVNVTDPVQPNPNQNLTSASV--CQKVTLAKWDAP 455
Db 1384 SPKCKVDVTNVTQPNPNQNLTAQAPNSMDAILKKNAP 1422

RESULT 5
I40229
C:Species: Porphyromonas gingivalis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40229
R:Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A:Title: Structural characterization of arginipain, a novel arginine-specific cysteine
A:Reference number: I40229; MUID:95168884; PMID:7864651
A:Accession: I40229

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-991 <RES>
A:Cross-references: GB:D26470; NID:G927644; PIDN:BAA05484.1; PID:G927645

Query Match 10.7%; Score 266; DB 2; Length 991;
Best Local Similarity 25.9%; Pred. No. 2e-10;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;

QY 30 TIDADGDNWTTTPPGTSGAGNSAICASSAYINPEGPQNDNVLVTELSLPPNGG 89
Db 595 SVNVCYNGAATATISANGKMF---GSAVVENGTATINLTG-----LTNES 637

QY 90 TLTFWVCAQADANYASEHVYASSTG--NDASNFANALLEVLTAKTVTAPAEARGTGV 147
Db 638 TLTLTV---VGYNKETVKTINTNGENPYQPVSN-----LTATT-----CGQKV 679

QY 148 QGTWYQKTVLPAGTKYVAFRHFCTDFFWINLDDV-EIKANGKRADETTFESSHGEA 206
Db 680 TLKWDAPSTKTNATTN-TARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWMDGS- 737

QY 207 PAEWTTIDADGGQGWCLSSGQLGWLTAHGTNVTVASFSWN-----GMLNPDNVLISK 261
Db 738 -GYQILLDADHDQYQGVIPSDTHLMPNCVPAFLFAPFEYTVPENADPPSCPTNMIM-- 794

QY 262 DVTGATKXYAYVNDGFGPDHYAVMISKTGTNAGDFTVFEETPNGINKKGARFGLSTE 321
Db 795 DGTASVNI-----PAGTY-----DFAI-----811

QY 322 ADGAKPOS---VWI-----ERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQTMTGSP 371
Db 812 ---AAPQANAKIWIAGQPTKEDDYVFEAGKKY----HFLMKMGSGDGTTELTISEGGG- 863

QY 372 TPTDVTYTVYRGTGKIKEGLTTFTEEDGATGNHEVCVEVKYTAGVSKPCVNTVDPV 431
Db 864 --SDITYTVYRGTGKIKEGLTETTYRDAGMSAQSHCYVEVKYAGVSPKVCVDYIPDV 921

QY 432 QFNPNQV-NLTGSAGVQKVTLL 450
Db 922 ADVTAQKPYTLTVGKTIIV 941

RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AF2959
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
sler, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF2959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1052 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:gl7741659; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3276
A:Map position: linear chromosome

Query Match 5.6%; Score 139.5; DB 2; Length 1052;
Best Local Similarity 20.9%; Pred. No. 0.12;
Matches 123; Conservative 67; Mismatches 159; Indels 239; Gaps 33;

QY 13 TTTLSESPENGIPASWKTIDADGNNWTT-----TPPGTSGAGNSAICASSAS 64
Db 52 TVTSGEALIGRHSASEATV--TGDGSKWTTDQVGQDTSPPGG--LAGNGT-----99

QY 167 FRHFGCTDFFWNL-----DDVEIKANG-----KRADFTETTFESSTHGEPAEAWTTTIDAGD 218
Db :
476 SR-----SD-----LNVGLYNGSGMTVEAGGAVKSRGYVATYGGST-----SAVVVTGD 520

QY 219 GOGMWCLSSQLGWLTAKGGT-NNVASFSNNGMALNPNDNYLI-----SKDVTCGA-TK 268
Db :
521 GGSMAWTGTFFVG--ASGATGNVTVS---NGGAIRATGVTGLDLAGASGTMITTGAGSK 575

QY 269 VKYYY-----AVNDGPPG-----DHV-----AMVTSKTGT--- 293
Db :
576 VTAYVDNGTVNSGVDFGQSGSLSVVNGSSLDAYNLYVGNALGSGGVAIVSGVSHVS 635

QY 294 -----NAGD-----FTVFBETPNINKGARFGLSTEADGAPQSV 330
Db :
636 VDGLMVVGNAGNSVEITCGASLASLAAPTILIIATEAGSTGVLISGAGSGQTARSAGA----- 690

QY 331 WIE-RTVDLPAGTKYVAFPHYNCSLDNIALLDDIQ-----FTMGSGSPTPDYTY 378
Db :
691 -VEARAIFAGAGNSIVFNH--SETGYTLISADISGAGRVAEAGVTTLISGNS----- 740

QY 379 TVYRDGTXIKBGLTETT-----FTVVBETPNINKGARFGLSTEADGAPQSV 402
Db :
741 --YSGGTTISAGMLKGTAKSFGSGGIVNNAELVDDGGTLLSNALSGTGSFEKTGDCNILL 798

QY 403 TGNHSYCVVEKYTAGVSPKECVNTVPQFPNPQNITGSAGVCQKYL 450
Db :
799 TGNSTYSGATAVSAG-----KLVSNGSLASAVSNGSATVGGTGTI 839

RESULT 8
B75622
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: B75622
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson,
M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski,
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans
A:Reference number: A75250; NCBI:20036896; PMID:10567266
A:Accession: B75622
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-691 <WHI>
A/Cross-references: GB:AB001826; NID:g6460827; PIDN:AAF12628.1; PID:g6460924; T:
C:Genetics:
A:Gene: DEB0037
A:Map position: megaplasmid
A:Genome: plasmid
A>Note: plasmid MP1

Query Match 5.3% Score 132.5; DB 2; Length 691;
Best Local Similarity 21.9%; Pred. No. 0.21;
March 11, 2001

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|---|---|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|
| QY | 4 | NNPNPNPGTTTTSESFENGIPASWKTIADADGNNWTTTP | 250 | | 255 | | 260 | | 265 | | 270 | | 275 | | 280 | | 285 | | 290 | | 295 | | 300 | | 305 | | 310 | | 315 | | 320 | | 325 | | 330 | | 335 | | 340 | | 345 | | 350 | | 355 | | 360 | | 365 | | 370 | | 375 | | 380 | | 385 | | 390 | | 395 | | 400 | | 405 | | 410 | | 415 | | 420 | | 425 | | 430 | | 435 | | 440 | | 445 | | 450 | | 455 | | 460 | | 465 | | 470 | | 475 | | 480 | | 485 | | 490 | | 495 | | 500 | | 505 | | 510 | | 515 | | 520 | | 525 | | 530 | | 535 | | 540 | | 545 | | 550 | | 555 | | 560 | | 565 | | 570 | | 575 | | 580 | | 585 | | 590 | | 595 | | 600 | | 605 | | 610 | | 615 | | 620 | | 625 | | 630 | | 635 | | 640 | | 645 | | 650 | | 655 | | 660 | | 665 | | 670 | | 675 | | 680 | | 685 | | 690 | | 695 | | 700 | | 705 | | 710 | | 715 | | 720 | | 725 | | 730 | | 735 | | 740 | | 745 | | 750 | | 755 | | 760 | | 765 | | 770 | | 775 | | 780 | | 785 | | 790 | | 795 | | 800 | | 805 | | 810 | | 815 | | 820 | | 825 | | 830 | | 835 | | 840 | | 845 | | 850 | | 855 | | 860 | | 865 | | 870 | | 875 | | 880 | | 885 | | 890 | | 895 | | 900 | | 905 | | 910 | | 915 | | 920 | | 925 | | 930 | | 935 | | 940 | | 945 | | 950 | | 955 | | 960 | | 965 | | 970 | | 975 | | 980 | | 985 | | 990 | | 995 | | 1000 | | 1005 | | 1010 | | 1015 | | 1020 | | 1025 | | 1030 | | 1035 | | 1040 | | 1045 | | 1050 | | 1055 | | 1060 | | 1065 | | 1070 | | 1075 | | 1080 | | 1085 | | 1090 | | 1095 | | 1100 | | 1105 | | 1110 | | 1115 | | 1120 | | 1125 | | 1130 | | 1135 | | 1140 | | 1145 | | 1150 | | 1155 | | 1160 | | 1165 | | 1170 | | 1175 | | 1180 | | 1185 | | 1190 | | 1195 | | 1200 | | 1205 | | 1210 | | 1215 | | 1220 | | 1225 | | 1230 | | 1235 | | 1240 | | 1245 | | 1250 | | 1255 | | 1260 | | 1265 | | 1270 | | 1275 | | 1280 | | 1285 | | 1290 | | 1295 | | 1300 | | 1305 | | 1310 | | 1315 | | 1320 | | 1325 | | 1330 | | 1335 | | 1340 | | 1345 | | 1350 | | 1355 | | 1360 | | 1365 | | 1370 | | 1375 | | 1380 | | 1385 | | 1390 | | 1395 | | 1400 | | 1405 | | 1410 | | 1415 | | 1420 | | 1425 | | 1430 | | 1435 | | 1440 | | 1445 | | 1450 | | 1455 | | 1460 | | 1465 | | 1470 | | 1475 | | 1480 | | 1485 | | 1490 | | 1495 | | 1500 | | 1505 | | 1510 | | 1515 | | 1520 | | 1525 | | 1530 | | 1535 | | 1540 | | 1545 | | 1550 | | 1555 | | 1560 |
|----|---|---|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|-----|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|--|------|

| | | |
|----|------|---|
| Db | 1044 | TPSTPLADGTVVNATDPAG-NTGGG---SITVDIAIAPAT-----FTVNLSNGSSLSG-1094 |
| Qy | 348 | RHNCSDNLVILDD-----IQFTMGSSPTDPTVYTVYRDGTKIKESLTETTFEEDGV401 |
| Db | 1095 | ---TABPGSVIILTDGNGNPIAEVTDGSG--GNWYTYT---PSTPIANGVFNVNVQAQD--1143 |
| Qy | 402 | ATGNHBYCYEVKTKTAGVSPKECVN-----VTVDVPQFNQVQLTGSAVG445 |
| Db | 1144 | ASGNSSGPPATVTVDDSGAPPAPVNPNGVVISGTBAGATVITLTDAGNPIIGOVATDGG1203 |

RESULT 10
C89874
autolysin
C/Species
C/Date: 1
C/Accessi
R/Kuroda,
ma, A.; M
C.; Shib
Lancet 35
A/Title: 1
A/Referen
A/Accessi
A/Status:
A/Molecul
A/Residue
A/Cross-r
A/Experim
C/Genetic
A/Gene: a

[illegible]

RESULT 11
AH2515
hypothetic
C:Species

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2515
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saamamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4936 <R>
A;Cross-references: GB:BA000020; PIDN:BA078388.1; PID:gl7135842; GSPDB:GN00180
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alx7304
A;Genome: plasmid

Query Match 5.2%; Score 128; DB 2; Length 4936;
Best Local Similarity 20.4%; Pred. No. 5.5;
Matches 108; Conservative 60; Mismatches 171; Indels 190; Gaps 27;
QY 1 GTPNPNNPNTG-----TTLSEFENGIPASWKIDADGCGNWT-----TPPPGG 48
Db 3161 GLSNFN-----TPGTQQLQVADATVQDNAGN-----SGDAARTTFTTAAAPPTFGV 3206
QY 49 TSPAGHNSAICASSASYINFRGPNQPNLYV-----TPELSLP-NGG-----TLTF 93
Db 3207 T-----ITQSGGSTAVIEG-GNTDSYTLVLRTOPTADVTVLNTGSGQITTDKTLTF 3257
QY 94 -----WVCAQ-----DANVASEHAYVASTGNDASNFANALLEEVLTAKTVVTAPEA 141
Db 3258 TSANWNTPQITVNAVNDITTEGNHTSTISHSISSTDTNYSNVTLPDI--AVSITDNDAE 3315
QY 142 IRGRVQGTWYQKTVQVLPAGIKYVAFRHFGCTDFWNLDDVEIKANGKRA--DFTETFE 199
Db 3316 IRGK-----WNDIDGGVDKDTGEPGLQGWITLYD 3345
QY 200 SSTHGEAP--AEWTTIDADGGQGLCLSSG-----QLGMLTAHGGTNTV----- 241
Db 3346 SNTNGOLDNGEISITTDANGNYQ--FTNLPGYVTVAEVQPGWKQTPPGTNTTINADIP 3404
QY 242 -----VASFSWNGWAIN--PNYLISKDVTGATKYYVAVNDGFPDGHYAVMISKT 291
Db 3405 LAIPSLDMISFGDSNGIQLNFSAAYIVKEDGTAITEV---WVTRTGNSTSAVSATLSFT 3461
QY 292 -GTNAG-----DFTWTFEE-----TPNGIN----- 310
Db 3462 DGTATGCGGASSVNDPNNVPPFIATFENETSKLISQVALLANPNAIKRNDKVEGN 3521
QY 311 -----KGGARFGLSTEA-----DGAKPQSVWIERTVDLPLAG--TKYVAFRHYNCS 353
Db 3522 EYFTIKLNTPTGGAVIGNQSIATVTIIDDEAPSDITVTPLETSTTITSADVQAIYLI 3591
QY 354 DLNVLIDDDTQFTMGSGPTDYTVYVRDGTCKLKEGLTETTFEEDGVA 402
Db 3592 NLNNFWADSRPANIKGN-----DFTSVIIDTGLNLNHPFFGADTNDNGIA 3626

RESULT 12
S10789
amylase A-180 - alkaliphilic eubacterium 163-26
C;Species: alkaliphilic eubacterium 163-26
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C;Accession: S10789
R;Candussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A;Title: Biochemical and genetic analysis of a maltotetraose-producing amylase from an alkaliphilic eubacterium 163-26
A;Reference number: S10789; MUID:90336627; PMID:1696201
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1684 <CAN>

A;Cross-references: EMBL:X53373; NID:g48305; PIDN:CAA37453.1; PID:g48306
Query Match 5.1%; Score 127.5; DB 2; Length 1684;
Best Local Similarity 20.9%; Pred. No. 1.5;
Matches 102; Conservative 53; Mismatches 184; Indels 149; Gaps 27;
QY 23 GIPASWKITDADGGNNTWTPPPGGTSTFAGHNSAICA-SSASYINFEFGPNQNDYLVTP 81
Db 204 GLPRDWTNCA--QGQNWHT-----HNDIMNKDNEAANWWSGD----- 241
QY 82 ELSLNGGTLTFWVCAOD--ANYASEHAYVASTGNDASNFANALLEEVLTAKTV--VTAP 139
Db 242 -----WIRADETAGY-----DNGGSEQTCIGFLPDIKTEVTITGVLDLP 280
QY 140 EAIRGT--RVQGTWYQKTVQVLPAGIKYVAFRHFGCTDF--WINLDDVEIKANGKRAFDTE 196
Db 281 PILRKNWDQASGYEDWF--VFAAEPRQDLNIAPKYLIRKITSWEEFIDGFRVDTAK 339
QY 197 TFSSTHGEAPAE-----WTTIDADGGQGLCLSSGQLGMLTA-----HG----- 237
Db 340 HVEIERWAELEKNEAEVALQOTWRENNPDKPGANW-----DDNFMTAEVFGHGLKSEYED 394
QY 238 -GNVVASFSW-----NGMALPNDNVLISKDVTGATKYYVAVNDGFP 280
Db 395 FGDSVINFEFQANFNFNLEGLFSRYANSINTDPDNMLSYVSSHDTKL---YSRDD--- 448
QY 281 GDHYAVMISKTGN---ACDFTVVF--EETPNGINKGARFGLSTEDAGKAPQSVW--IE 333
Db 449 -----LIQAGTALLLLPGQVQVFGDETARPLDGGG-----SDPEQGTSSNMNANIN 496
QY 334 RTVDLPAGTKYVAFR-----HYNCSDLNYI-----LLDDIOFTMGGSPTPT 374
Db 497 QNV--LSHWQKLGQFRNNHIAIGAGAHQKLSDSPTFARTVYESDDIDVDEVVVATGAQGTTA 555
QY 375 DYTYYVVRDGTCKLKEGLT--ETTDEGVTAGTNGHEVCVEKYTAGVSPKCVNVTVDPVQ 432
Db 556 VTVEGVFEDGTGVRDAYDTGDETTVTK--GTAT-----FTAGTQGIILIENTAPPT 604
QY 433 FNPVQNL 440
Db 605 NLPIVSAT 612
RESULT 13
B75489
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: B75489
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mc
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <WHI>
A;Cross-references: GB:AE001925; GB:AE000513; NID:g6458383; PIDN:AAF10267.1; PID:g64583;
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0685
A;Map position: 1
Query Match 5.1%; Score 127; DB 2; Length 713;
Best Local Similarity 21.8%; Pred. No. 0.51;
Matches 129; Conservative 47; Mismatches 214; Indels 208; Gaps 29;
QY 2 TPNPFP-NPNPGTTLSE-----SFENGIPASWKITDADGGNNTWTPP---PGGT--- 49
Db 86 TPSTPTNPNPSETIETFPVTVVNPVPSFTIT---PNDGSKDTTQPTDYNFQGTREV 142

QY 50 -----SFAGHNSAICASSASYINFE-----GFQNPNDYLVTPPE-----LSLNGGTL-- 91
Db 143 PCDKNVTF-----AYLTNTGNVNGESYTLTNPDPFGAVKTPENTIRFYLDSSNNQGLDQ 197
QY 92 -----TFWVCAQDANYASEHVAVY-----ASTGNDASNFAN--ALLE 127
Db 198 SEIAAGATTITDVAINQTVKFFQVYDVPCTATSTDKFGGDPGTGRNDNPNFSDNPTLPR 257
QY 128 EVLTAKTV-----TAPEAIRGTRVQGWYQKTV--QLPA 160
Db 258 DANNSTVTINRKDGVVIGPKADPDGNGNPVTPAYNSPEGINVPTASDTQVATVTLPT 317
QY 161 GTKYVAFRHFPGCTDFEWINLDDVEIKANGKRADEFTEFSSSTHGEAPAEWTTIDADGQ 220
Db 318 SGVTVTFTN-----TIQNTGNR-----TDTFELTQNTTPAGTTVVPKANGN 360
QY 221 GMLCLSSGQLGMLTANGGTTNVASFWSNGMALPNPNLYLSKDVGTAKVKYVAVNDGFP 280
Db 361 A-----LPTVNGKPVQNVPENGTV-----DIQVITLPAVTPQTOL-----SGQP 401
QY 281 GDHYAVMISKGTGN-----AGDFT-----VVFEEPTNGINKGARFGLSTADGAKPOSWIERT 335
Db 402 -----AVTITTSQNDPTKSDTTKDIIEVKVP-----GIAFGDPTPLGGLDPTPVGTPT 451
QY 336 -----VDLPAG-----TKYVAFRHNCSIDLNYILLDDIOFTMGG-----SPT 372
Db 452 GVPNGFTPLTPGNPQCTAPIRTYLPMELIANLGSQDDAFVVSGTAPVTVLPNDGTVNPT 511
QY 373 PTDYTVTVRD-----GKIKRGLTETTFEEDGVATGNHEVCVEVKY-----414
Db 512 PVIVPVVYRYRDNVNGDKGLDAGDTL-----QCGNTGTIKPGEELKLIADVVDVPCAAAQQT 567
QY 415 -----TAGVSPKE-----CVNVTVDVPQFNPVQNLTCGSAVGQKVT 449
Db 568 TLNQEAKSPPTGVQSQDPNDITTVGGNGKPIVTKTVDRKATANPGDPLTYITIGKNTS 625

RESULT 14
A86685
prophage pil protein 46, tail component [imported] - Lactococcus lactis subsp. lactis (s
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: A86685
Genome Res. 11, 731-753, 2001
R: Bolotin, A.; Wincker, P.; Mauer, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrli
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: A86685
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:Cross-references: GB:AE005176; PID:gl2723361; PIDN:AAK04579.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: pil46

Query Match 5.1%; Score 126; DB 2; Length 1441;
Best Local Similarity 22.3%; Pred. No. 1.5;
Matches 98; Conservative 49; Mismatches 133; Indels 160; Gaps 26;
QY 30 TIDADGNNWTTTPPGGTSFAGHNSAICASSASYINFEQPNP-----DNYLVTPELSL 85
Db 602 TISSG-----TVPTTGTS-----QVPLVKEQYL-----628
QY 86 PNGGTLTFWVCAQDANYASEHVAV--YASSTGN-----DASNPANALLEVLTAKTV 135
Db 629 ----WTKLTWYTD--NTSTGVSVSVAIXDGNNGHGDGFPKDGVGISNTIIEV-----677
QY 136 VTAPAIRGCT-RVQGWYQKTVQLPAGTKYVAFRHF-----GCTDFWINL-----180
Db 678 ----GAVSTSRPTGWSITITVPAG-QYLWTRTTWQYTDGTEQGYINALMGLTGASG 732
QY 181 -DDVEIKANGKRADEFTE--TFESSHGE--EAPAE-----WTTIDAG 217

Db 733 RDGIAGK-DKGIKATAITTYQASTNGTTTAPTGTWSTSVPSVAKGSFLWTRTIWYTD--- 788
QY 218 DGQWCLSSGQLGMLTANGGTTNVASFWSNGMALPNPNLYLSKDVGTAKVKYVAVN- 276
Db 789 -----NTTETGYAVAYMGIN-----GNGTNGIAGCKDGTGKTTITVAVGT 830
QY 277 -----DGFPGDHYAVMISKGTNAGDPTVVFEEPTNGINKGARFGLSTEA 322
Db 831 SGTAPTGGWNSQVNPVPAQY--LWTKT-----VWDYTDKTSYTSVSKFGEK-GDKGQ 884
QY 323 DGAKPQSVWIERVDLPAG-----TKYVAFRHNCS--LNYILLDD-----IOFTMG 368
Db 885 GVQGIQGVDRGQIPGPKGADKTIQTHIAYANSADGKTNFTSDSNRTYIGMYVDENIN 944
QY 369 GSPTPTDYTYTVR--DGTK 386
Db 945 DSTTSPDSYSLVKGADGTQ 964

RESULT 15
C86822
hypothetical protein yqpk [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86822
Genome Res. 11, 731-753, 2001
R: Bolotin, A.; Wincker, P.; Mauer, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrli
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1649 <STO>
A:Cross-references: GB:AE005176; PID:gl2724583; PIDN:AAK05677.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yqpk

Query Match 5.1%; Score 126; DB 2; Length 1649;
Best Local Similarity 20.4%; Pred. No. 1.8;
Matches 110; Conservative 56; Mismatches 189; Indels 184; Gaps 28;
QY 46 PGGTTFAGHNSAICASSASYINFEQ-PQN-----PDNY-----LVTPELSLEN 87
Db 65 PGGTTFNG-----NSVSQPNRTLPQNSVEDAIMNSFGWQQTIANLTLPNQVTAQV 117
QY 88 GGTTLTFWVCAQDANYASEHVAVYASSTGNDAGNF-----ANALLEVLTAKTWTAPEA- 141
Db 118 NGTIIGSYPTTDAGVQQAALYMY-SATNASTSDFGLYYGANTLNSL-LKAVSNPSAT 175
QY 142 -----IRGTRVQGWYQKTVOL-----PAGTKYVAFR-----HFGCTDFWINLDDV 183
Db 176 NMTFTSLKHAHAKSLTWISNPADLLTSSNSQSPSGNTYGTLPNSVYFGVPTTF----RVN 231
QY 184 EIKANGKRADEFTEFESSHGEAPAEWTT-----IDADGQGW 222
Db 232 TVRAGSDNIIYAOGNAPATNG-----SWITGAPNIYGGTDSISGNTNLIYGATSIAGW 287
QY 223 LCLSSGQLGMLTANGGTTNV-----VASFSWNGMALPNPNLYLSKDVGTG--ATKV 269
Db 288 -NIYGNASAATISGNTHVTLAQSSSTINSVTGSGASGTTISGNTNL-----DISGAIASQI 343
QY 270 KYVAVNDGFPDHYAVMISKGTNAGDPTVVFEEPTNG-----INKGARFG-----LS 319
Db 344 TNIYGAGIG-----TSNSPVNVNGNVITVYVNSTNGGARVQLYQGGTVVGNISGSY 394
QY 320 TEADCAKQSVWIERVTDL-----PA-----GKTVAFRHNCSDLNVL 359
Db 395 NTLGSA-----GGWTGATSNINGAGGPASTFNNGSGFOGNTGTSGAGNVISNYSNTSFT--- 448
QY 360 LDDIOFTMGGSPPTDYTYTVRDRDCKIKRGLTETTFEEDGVATGNHEVCVEVKYTAG-- 417

Db 449 TGOALFTGNGAGTASYAQA-----TNSTTAAQGILYANITNYIKSAFTTGTA 496
Qy 418 -----VSPKE-----CVNVTVDVPQENPVPONLTGSAVGOKVT 449
Db 497 GAVYGIVCGNGHDSLKISPSQWGLGSGTGLDSAVGVT-DAKAYGQIPSTTVVSNQAQIT 554

Search completed: May 18, 2004, 11:47:50
Job time : 12.906 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 8.3342 Seconds
(without alignments)

2848.981 Million cell updates/sec

Title: US-08-570-311-20

Perfect score: 2480

Sequence: 1 GTPNPNPNPNTTILSESF.....QNLTSAGVGKVLKWDAPN 456

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 2480 | 100.0 | 2628 | 1 HGA2_PORGI | Q51845 porphyromon |
| 2 | 2441 | 98.4 | 2164 | 1 HGA2_PORGI | P59915 porphyromon |
| 3 | 1453 | 58.6 | 989 | 1 PRTH_PORGI | P46071 porphyromon |
| 4 | 2266 | 10.7 | 991 | 1 CPFG_PORGI | P28784 porphyromon |
| 5 | 131.5 | 5.3 | 1256 | 1 ATL_STRAU | P52081 staphylococ |
| 6 | 126 | 5.1 | 721 | 1 OGP_MOUSE | Q62010 mus musculu |
| 7 | 121.5 | 4.9 | 465 | 1 SLAP_LACBR | Q05044 lactobacilli |
| 8 | 121 | 4.9 | 1090 | 1 GUXB_CELFI | P50899 cellulomona |
| 9 | 118.5 | 4.8 | 1385 | 1 C5AA_BACUD | Q45760 bacillus th |
| 10 | 114 | 4.6 | 872 | 1 GUXA_CELFI | P50401 cellulomona |
| 11 | 113 | 4.6 | 2660 | 1 YEEU_ECO57 | Q8x8v7 escherichia |
| 12 | 112.5 | 4.5 | 1034 | 1 BGAL_BACME | O52847 bacillus me |
| 13 | 112.5 | 4.5 | 1260 | 1 ALS1_CANAL | P46590 candida alb |
| 14 | 110.5 | 4.5 | 1122 | 1 ADPI_MYCGA | Q49379 mycoplasma |
| 15 | 109.5 | 4.4 | 331 | 1 PME_ASPAC | Q12535 aspergillus |
| 16 | 109.5 | 4.4 | 607 | 1 YK28_ARCFU | O28251 archaeroglob |
| 17 | 108.5 | 4.4 | 1953 | 1 BIGA_SALTY | P25927 salmonella |
| 18 | 108 | 4.4 | 699 | 1 CHII_BACCI | P20533 bacillus ci |
| 19 | 108 | 4.4 | 987 | 1 YD94_METJA | Q58789 methanococ |
| 20 | 108 | 4.4 | 1289 | 1 C5AB_BACUD | Q45753 bacillus th |
| 21 | 107.5 | 4.3 | 551 | 1 AMVB_THETU | P19584 thermococ |
| 22 | 107.5 | 4.3 | 1200 | 1 HYAL_STRPU | O76536 strongyloce |
| 23 | 107 | 4.3 | 1656 | 1 OMPB_RICJA | O06653 r outer mem |
| 24 | 107 | 4.3 | 2249 | 1 OMPA_RICRI | P15921 rickettsia |
| 25 | 106.5 | 4.3 | 1122 | 1 ADP2_MYCGA | Q9rem8 mycoplasma |
| 26 | 106.5 | 4.3 | 2124 | 1 PGCA_RAT | P07897 rattus norv |
| 27 | 106.5 | 4.3 | 2358 | 1 YEEU_ECOLI | P76347 escherichia |
| 28 | 106 | 4.3 | 1045 | 1 PRTH_SERMA | P09489 serratia ma |
| 29 | 106 | 4.3 | 1300 | 1 L20K_RICRI | P14914 rickettsia |
| 30 | 106 | 4.3 | 1654 | 1 OMPB_RICRI | Q53047 r outer mem |
| 31 | 105.5 | 4.3 | 1220 | 1 C5AC_BACTO | P56955 bacillus th |
| 32 | 105.5 | 4.3 | 1534 | 1 YFAS_ECO57 | Q8xe35 escherichia |
| 33 | 105 | 4.2 | 827 | 1 XANP_XANS2 | Q60106 xanthomonas |

| | | | | | |
|----|-------|-----|------|--------------|---------------------|
| 34 | 105 | 4.2 | 1045 | 1 GUNB_CELFI | P26225 cellulomona |
| 35 | 105 | 4.2 | 2812 | 1 ZAN_HUMAN | Q9Y493 homo sapien |
| 36 | 103.5 | 4.2 | 837 | 1 XYNZ_CLOTM | P10478 clostridium |
| 37 | 103.5 | 4.2 | 1655 | 1 OMPB_RICCN | Q9kka3 r outer mem |
| 38 | 103.5 | 4.2 | 1861 | 1 APU_THETU | P38536 t amylopull |
| 39 | 103 | 4.2 | 1157 | 1 C8AA_BACUK | P245704 bacillus th |
| 40 | 102.5 | 4.1 | 282 | 1 PRTH_ASPNG | P24665 aspergillus |
| 41 | 102.5 | 4.1 | 1045 | 1 PRTH_SERMA | P29805 serratia ma |
| 42 | 102 | 4.1 | 435 | 1 AM3D_ORYSA | P27933 oryza sativ |
| 43 | 102 | 4.1 | 941 | 1 GUN_BACS6 | P19424 bacillus sp |
| 44 | 101.5 | 4.1 | 1462 | 1 PTP6_DROME | P16620 drosophila |
| 45 | 101.5 | 4.1 | 2132 | 1 PGCA_MOUSE | Q61282 mus musculu |

ALIGNMENTS

RESULT 1
HGA2_PORGI
ID HGA2_PORGI STANDARD; PRT; 2628 AA.
AC Q51845;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin A precursor.
GN HAGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_taxid=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97047672; PubMed=8926061;
RA Han N., Whitlock J., Progulski-Fox A.;
RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381
contains four large, contiguous, direct repeats.";
RL Infect. Immun. 64:4000-4007(1996).
CC -!- FUNCTION: Agglutinates erythrocytes.
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41807; AAB17128.1; -.
DR FIR; T28651; T28651.
KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;

Query Match 100.0%; Score 2480; DB 1; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.6e-163;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPNTTILSESFENGIPASWKTIADGGNNWTTTPPGGTSFAGHNSAICA 60
Db 950 GTPNPNPNPNTTILSESFENGIPASWKTIADGGNNWTTTPPGGTSFAGHNSAICA 1009
QY 61 SSASVINFEFGPNPNLVTPELSLPNGGTLTFWCAQDANYASEHYAVYASTGNDASN 120
Db 1010 SSASVINFEFGPNPNLVTPELSLPNGGTLTFWCAQDANYASEHYAVYASTGNDASN 1069

QY 121 PANALLEEVLTAKTVTVAPEAIRGTRVGQTKVQKTVQVLPAGTKYVAFRHFCTDFFWINL 180
 DB 1070 FANALLEEVLTAKTVTVAPEAIRGTRVGQTKVQKTVQVLPAGTKYVAFRHFCTDFFWINL 1129
 QY 181 DDVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTTN 240
 DB 1130 DDVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTTN 1189
 QY 241 VVASFSGMALPNPNLYLSKDVATGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTV 300
 DB 1190 VVASFSGMALPNPNLYLSKDVATGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTV 1249
 QY 301 VFEETPNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNLYLL 360
 DB 1250 VFEETPNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNLYLL 1309
 QY 361 DDQFTMGSGPTPTDYTYTVYRDGPKIKEGLTETTFEEDGVATGNHGYCUEVKYTAGVSP 420
 DB 1310 DDQFTMGSGPTPTDYTYTVYRDGPKIKEGLTETTFEEDGVATGNHGYCUEVKYTAGVSP 1369
 QY 421 KECVNVTVDPQFNPVQNLITGSVAGQKVTILKWDAPN 456
 DB 1370 KECVNVTVDPQFNPVQNLITGSVAGQKVTILKWDAPN 1405

RESULT 2

HGAL_PORGI
 ID HGAL_PORGI STANDARD; PRT; 2164 AA.
 AC P59915;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hemagglutinin A precursor.
 GN HAGA OR PG1837.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RA MEDLINE=22829867; PubMed=12949112;
 RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
 RA Granger D., Tetzelin H., Dong H., Galvin J.L., Duncan M.J.,
 RA Dewhirst F.E., Fraser C.M.;
 RT "Complete genome sequence of the oral pathogenic bacterium
 RT Porphyromonas gingivalis strain W83";
 RL J. Bacteriol. 185:5591-5601(2003).
 CC -!- FUNCTION: Agglutinates erythrocytes (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE017178; AAQ66831.1; AUT_INIT.
 DR TIGR; PG1837; -.
 DR
 KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL
 FT CHAIN 26 2164 HEMAGGLUTININ A.
 FT DOMAIN 26 539 PEPTIDASE C25-LIKE 1.
 FT DOMAIN 540 991 PEPTIDASE C25-LIKE 2.
 FT DOMAIN 992 1443 PEPTIDASE C25-LIKE 3.
 SQ SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;

Query Match

98.4%; Score 2441; DB 1; Length 2164;

Best Local Similarity 98.9%; Pred. No. 6.2e-161;
 Matches 449; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 3 PNPENPNPGTTTLLSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAICASS 62
 DB 496 PNGTNPENPGTTTLLSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAICVSS 555
 QY 63 ASYNFEPGPNPNLYLTPVLSLNGGTLTFWVCAQDANYASBHYAVASSTGNDASFA 122
 DB 556 ASYNFEPGPNPNLYLTPVLSLNGGTLTFWVCAQDANYASBHYAVASSTGNDASFA 615
 QY 123 NALLEEVLTAKTVTVAPEAIRGTRVGQTKVQKTVQVLPAGTKYVAFRHFCTDFFWINLDD 182
 DB 616 NALLEEVLTAKTVTVAPEAIRGTRVGQTKVQKTVQVLPAGTKYVAFRHFCTDFFWINLDD 675
 QY 183 VEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTTN 242
 DB 676 VEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTTN 735
 QY 243 ASFSWNGMALPNPNLYLSKDVATGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTV 302
 DB 736 ASFSWNGMALPNPNLYLSKDVATGATKVKYVAVNDGFGPDGHYAVMISKTGTNAGDFTV 795
 QY 303 EETPNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNLYLLDD 362
 DB 796 EETPNGINKGARGFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNLYLLDD 855
 QY 363 IQFTMGSGPTPTDYTYTVYRDGPKIKEGLTETTFEEDGVATGNHGYCUEVKYTAGVSPKE 422
 DB 856 IQFTMGSGPTPTDYTYTVYRDGPKIKEGLTETTFEEDGVATGNHGYCUEVKYTAGVSPKE 915
 QY 423 CVNVTVDPQFNPVQNLITGSVAGQKVTILKWDAPN 456
 DB 916 CVNVTVDPQFNPVQNLITGSVAGQKVTILKWDAPN 949
 RESULT 3
 PRTH_PORGI
 ID PRTH_PORGI STANDARD; PRT; 989 AA.
 AC P46071;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protease prth (SC 3.4.22.-).
 GN PRTH.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RA MEDLINE=95012612; PubMed=7927685;
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RT "Cloning and characterization of a new protease gene (prth) from
 RT Porphyromonas gingivalis";
 RL Infect. Immun. 62:4279-4286(1994).
 RN [2]
 RP ERATUM.
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RL Infect. Immun. 62:5707-5707(1994).
 CC -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
 CC P-GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
 CC IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
 CC AND IS A VIRULENCE FACTOR.
 CC -!- SUBCELLULAR LOCATION: In membrane vesicles.
 CC -!- SIMILARITY: Belongs to peptidase family C25.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

entireties requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L27483; AAA51298.1; --
HSSP; P23882; 1FWT.
MEROPS; C25_001; --
InterPro; IPR002376; formyl transf.
Pfam; PF00551; formyl_transf; 1.
KW Hydrolase; Thiol protease; Repeat; Virulence.
FT REPEAT 270 323
FT REPEAT 528 581
SQ SEQUENCE 989 AA; 110238 MW; FA85FE8A3AC8944C CRC64;

Query Match 58.6%; Score 1453; DB 1; Length 989;
Best Local Similarity 84.9%; Pred. No. 5.7e-93;
Matches 276; Conservative 11; Mismatches 30; Indels 8; Gaps 3;
136 VTAPAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFG-----CTDPWINLDDVEIKANGKR 191
34 VRSPEARIG-RIQGTWRQKTVLDPAGTEICFPSPLPKAPICTSTLMEL-----RSKTNKR 89
192 ADFTETPSSSTHGEAPAEWTTIDADGGQGLCLSSGQGLWLTARHGGTNVVASFSWNGMA 251
90 ADFTETPSSSTHGEAPAEWTTIDADGGQGLCLSSGQGLWLTARHGGTNVVASFSWNGMA 149
252 LNPNDYLISKDVTGATKYKYAVNDGPPGHYAVMISKTGTNAGDFTVVFETPNGINK 311
150 LNPNDYLISKDVTGATKYKYAVNDGPPGHYAVMISKTGTNAGDFTVVFETPNGINK 209
312 GGARFGLSTADGAKPQSWLERTVLDLPAGTKYVAFRHVNCSDLYILLDDIQFTMGSSP 371
210 GGARFGLSTADGAKPQSWLERTVLDLPAGTKYVAFRHVNCSDLYILLDDIQFTMGSSP 269
372 TPTDYYTYVYRDGTRKIEGLTETTFEEDGVTAGNHEYCVKVTAGVSPKECVNVTVDPV 431
270 TPTDYYTYVYRDGTRKIEGLTETTFEEDGVTAGNHEYCVKVTAGVSPKCVNVTVNST 329
432 QFNVPQNLGSAVGQKVTWKDAPN 456
330 QFNVPQNLKAQPDGDDVVLKWEAPS 354

RESULT 4

CPGL_PORGI STANDARD; PRT; 991 AA.
AC P28784; Q45168;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE (RGF-1).
GN RGFA OR RGF1.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 228-290 AND 517-541.
RC STRAIN=381;
RX MEDLINE=95168884; PubMed=7864651;
RA Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA Ikehara Y.;
RA "Structural characterization of argingipain, a novel
RT arginine-specific cysteine proteinase as a major periodontal
RT pathogenic factor from Porphyromonas gingivalis";
RL Arch. Biochem. Biophys. 316:917-925(1995).
RN [2]
RP SEQUENCE OF 228-270.
RC STRAIN=HG66;
RX MEDLINE=92406812; PubMed=1527017;
RA Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RA "Purification and characterization of a 50-kDa cysteine proteinase
RT (gingipain) from Porphyromonas gingivalis";

J. Biol. Chem. 267:18896-18901(1992).
-!- FUNCTION: Thiol protease which is believed to participate in intracellular degradation and turnover of proteins. Its proteolytic activity is a major factor in both periodontal tissue destruction and in bacterial host defense mechanisms. Activates complement C3 and C5.
-!- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and immunoglobulins, with a preference for Arg in P1, and hydrophobic residues in P2 and P3.
-!- ENZYME REGULATION: Requires cysteine for activation and Ca(2+) and/or Mg(2+) for stabilization. It is stimulated by glycine-containing dipeptides. It is resistant to inhibition by proteinase inhibitors in human plasma.
-!- SIMILARITY: Belongs to peptidase family C25.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D26470; BAA05484.1; --
PIR; I40223; I40229.
HSSP; P95493; 1CVR.
MEROPS; C25_001; --
InterPro; IPR007110; Ig-like.
InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
KW Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT SIGNAL 1 24
FT PROPEP 25 227
FT CHAIN 228 991
FT ACT_SITE 438 438
FT ACT_SITE 471 471
FT ACT_SITE 264 265
FT CONFLICT RT -> TK (IN REF. 2).
SQ SEQUENCE 991 AA; 108782 MW; 03EE3F43CEBE2544 CRC64;

Query Match 10.7%; Score 266; DB 1; Length 991;

Best Local Similarity 25.9%; Pred. No. 8.6e-11;
Matches 114; Conservative 50; Mismatches 164; Indels 112; Gaps 19;
30 TIDADGDNWTTTTPPGGTSFAGHNSAICASSAYINFEGPQNDYLVTFELSLKGG 89
595 SVNVCYDNGAIAITISANGKMF---GSAVVENGTATINLTG-----LTNES 637
90 TLTFWCAQDANYASEHYAVYASSTG--NDASNANALLEEVLTAQVVTAPAIRGTRV 147
638 TLTLTV-----VGYNKETVIKTINTNGBNPYQVSN-----LTATT-----QGQV 679
148 QGTWYQKTVOLPAGTKYVAFRHFGCTDFFWNLDDV-EIKANGRADFTETFEESTHGEA 206
580 TLKWDASTKINATN-TARSVDGIRELVLSVSDAPELLRSQAEIVLEAHVDWNGS- 737
207 PAEWTTTIDADGGQGLCLSSGQGLWLTARHGGTNVVASFSWN-----GMLNPDNYLISK 261
738 -GYQILLDADHDQYGVIPSDTHLWNCSPANLAPFEVTPENADPSCSPTNMIM-- 794
262 DVTGATKVKYYAVNDGPPGHYAVMISKTGTNAGDFTVVFETPNGINKGGARFGLSTE 321
795 DGTASVNI-----PAGTY-----DFAI----- 811
322 ADGAKPOS---VWI-----ERTVDLPAGTKYVAFRHVNCSDLYILLDDIQFTMGSSP 371
812 ---AAPQANAKIWIAGQGPTEKDDYVFEAGKY-----HFLMKMGSGDGETLTISEGG- 863
372 TPTDYYTYVYRDGTRKIEGLTETTFEEDGVTAGNHEYCVKVTAGVSPKECVNVTVDPV 431
864 --SDYTYTVYRDGTRKIEGLTETTYRDAGMSAQSHCYCVKVAAGVSPKCVDPVDPGV 921

QY 432 QFNPNVN-LTGSVGVQKVTL 450
 Db 922 ADVTAQKPYTLTVGKTTIV 941

RESULT 5
 ATL-STAAU
 ID ATL STAAU STANDARD; PRT; 1256 AA.
 AC P52081;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bifunctional autolysin precursor [includes: N-acetylmuramoyl-L-alanine
 DE amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
 DE acetylglucosamidase (EC 3.2.1.96)].
 GN ATL.

OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RP [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
 RC STRAIN=RM450;
 RX MEDLINE=95116542; PubMed=7816834;
 RA Oshida T., Sugai M., Komatsuza H., Hong Y.-M., Suginata H.,
 RA Tomasz A.;
 RA "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
 RT alanine amidase domain and an endo-beta-N-acetylglucosaminidase
 RT domain: cloning, sequence analysis, and characterization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=NCIC 8325-4;
 RC Foster S.O.;
 RA Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RL -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN
 CC HIGH-MANNOSE GLYCOPROTEINS AND GLYCOPROTEINS CONTAINING THE
 CC -[(MAN)5(GLCNAC)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
 CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
 CC OLIGOSACCHARIDE IS RELEASED INTACT.
 CC -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
 CC residues and L-amino acid residues in certain bacterial cell-wall
 CC glycopeptides.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
 CC unit in high-mannose glycopeptides and glycoproteins containing
 CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
 CC residue remains attached to the protein; the rest of the
 CC oligosaccharide is released intact.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
 CC EXTRACELLULAR LYTIC ENZYMES.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
 CC ACETYL-MURAMOYL-L-ALANINE AMIDASE FAMILY 2.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
 CC GLYCOSYL HYDROLASES.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D17366; BAA04185.1; -
 CC EMBL; L41499; AAA99982.1; -
 CC InterPro; IPR002502; Amidase_2.
 CC InterPro; IPR002901; Amidase_4.
 CC Pfam; PF01510; Amidase_2; 1.
 CC Pfam; PF01832; Amidase_4; 1.
 CC SMART; SM00644; Am_2; 1.
 CC SMART; SM00047; LY22; 1.
 CC Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
 CC SIGNAL 1 29
 CC POTENTIAL.

FT CHAIN 30 1256 BIFUNCTIONAL AUTOLYSIN.
 FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
 FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
 FT REPEAT 425 589 1.
 FT REPEAT 596 758 2.
 FT REPEAT 770 932 3.
 SQ SEQUENCE 1256 AA; 137384 NW; 2BB76CAA292FDD20 CRC64;

Query Match 5.3%; Score 131.5; DB 1; Length 1256;
 Best Local Similarity 22.3%; Pred. No. 0.24; Indels 153; Gaps 26;
 Matches 105; Conservative 47; Mismatches 165;

QY 3 PNPENPNPGTTT-----LSSEFENGIPASWKTIDADGDG---NNWTTTPPP-----46
 Db 594 PTPTPTPKPTPTTNNKLTVSSLGV---AQINAKNGLFTTVYDKTKEVQKTPA 649

QY 47 -----GCTSF---AGHNSA-----ICASSASYINPEGPON-PDNYLVTP-----LSLPN 87
 Db 650 VTKEASLGKNGKFLVLDYNSPTLIGWKQGDVYNNNAKSPVVMQTYTVKPGTKLYSPVW 709

QY 88 G-----GTLTFWVCADANVASEHYAVYASSTGNDASNEAN-----ALLE 127
 Db 710 GTYKQEAGVSGTGNQTFKATKQOIDS-----IYLFETVNGSGWVSKAYLAVPAAPKX 765

QY 128 EVLTAKT-----VYTAPEAIR-----GTRVQGTWYQKTVQLPAGTKYVAFRRHF 171
 Db 766 AVAQPKTAVKAYTVTKPTQTTSKIAQVKNPNTGIRASVYEKTAK--NGAKY-ADRTFY 822

QY 172 CT-----DFFWNLDDVEIKANGKRADPTE---TFSSSTHGEAPAE 209
 Db 823 VTKEAHGNETYVLLNNTSHNIPLGWFWNVKDLNVQNLGKVKTKQKYTVNKSNNGLSNVP 882

QY 210 WTTIDADGDGQGWLCSSGOLGWLTAHGGTNVVASRSMN-----GVALNPDNYLVLSKDV 263
 Db 883 WGT-----KNQVILTGNI-----AQGTNATKQVSQKDVLYGTINNRTGMVNAKDL 931

QY 264 TGATKVK-----YYAVNDGPGDGHYAVMISKTGT-----NAGDFTVVVFETPN 307
 Db 932 TAPTAVKPTTSAAKDNYTYVIKNG-NGYVYVTPNSDTAKYSLKAFNEQFAVVKEQVIN 990

QY 308 G-----INKGARFGLSTEADGAK-----PQSVWIERTVLDLPACTKY 344
 Db 991 GQTYWYKLSNGKLAWKST--DLAKELIKYNTQGTMTLNQVAQIQAGLOY 1038

RESULT 6
 OGP MOUSE
 ID OGP MOUSE STANDARD; PRT; 721 AA.
 AC Q82010;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
 DE (Oviductin) (Estrogen-dependent oviduct protein).
 GN OVGPI OR OGP OR CHIT5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Oviduct;
 RX MEDLINE=96115001; PubMed=7492680;
 RA Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,
 RA Azaki Y.;
 RA "Molecular cloning and characterization of a mouse oviduct-specific
 RT glycoprotein.";
 RL Biol. Reprod. 53:285-294 (1995).
 CC -!- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
 CC in the fertilization process and/or early embryonic development.
 CC -!- SUBCELLULAR LOCATION: Secretory granules.
 CC -!- TISSUE SPECIFICITY: Epithelial cells of the oviduct.
 CC -!- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D32137; BAA06863.1; --
DR MGDI; MGI:106661; OvGP1.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE 18; FALSE NEG.
DR Glycoprotein; Fertilization; Repeat; Signal.
FT SIGNAL 1 21
FT DOMAIN 22 721
FT CHAIN 486 632
FT CARBOHYD 402 402
FT CARBOHYD 442 442
FT CARBOHYD 469 469
FT CARBOHYD 721 721
SQ SEQUENCE 721 AA; 78807 MW; 37246C8F01663652 CRC64;

Query Match
Best Local Similarity 5.1%; Score 126; DB 1; Length 721;
Matches 103; Conservative 55; Mismatches 185; Indels 202; Gaps 22;

QY 51 FAGHNSATCA-----SSASYINF----- 68
DB 222 FTGHSPLFLSPEDSKSAYAMNYRKLGTLPADKLIMGFTYGRNFYLLKESKNGLOTAS 281
QY 69 EGPQPNPNVLTPELSLPGNGTLTWVCAQDANYASEHYAVYASSTGNDASNFANALLEE 128
DB 282 MGPAFPGK-----TQAGFLATIEVCSFVQRAKKHWIDYQ----- 317
QY 129 VLTAKTVTAPAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFEGCTDFFWINDDDV----- 183
DB 318 -----VVPVAFKKEWLG--YDDTISFSYKAMYVKREHFGGAMVWTLDDMDVDRGTFC 367
QY 184 -----EIKANGKRAD-----FTETFSST----- 202
DB 368 GNGPPLVHLNELLVQTESNSTPLPQFWFTSSVNASGPGSENTALTEVLTDTIKILPP 427
QY 203 HGEAPA-----EWTTIDADGCGWGLCLSSGQLGWLTAHGTT-----NVVAS----- 244
DB 428 GGEAMTEVHRRYENWTTVPDGS-----VTPGTSAPRKHAVTPENNTWAAEAKTWTST 481
QY 245 ---FSWNGMALPNPNYLISKDVTAIKV-----KYIYAVNDGFPDGHYAVM-ISKGTGNA 295
DB 482 LDFESKTTTGVSKTTTGISKTTTGVSKTTTGVSKATAGISKITPEISKATAGVSKTTTGV 541
QY 236 GDFTVVFEETNGINKGARFGLSTEADGAKPQSVWIER-TVDLPAGTKYVAFRHYNCSD 354
DB 542 SKTTTGISKTTTGVSK--TTTGISKTTTGISKTTTGVSKTTTGVSKTTTGISKTTTGIS- 598
QY 355 LNYILLDDIQTWGGSPPTDYTYVYVRDGTIKIKEGLTETT-----FEEDGVAT 403
DB 599 -----QITTGISKTTTDISKTT-TGISKTTTGISKTTTGMVIVQQTQANEATTAT 648
QY 404 GNHEVCYEVK-----YTAGVSPKE-----CVNVTVDPVQFNPQNLJTGSAVGQK 447
DB 649 MDHQSVTPTMTDITLFLKLTMTPTSEKTSRKKTMVLEKATVSPREMSATPN-----GQS 702
QY 448 VTLKW 452
DB 703 KTLKW 707

RESULT 7
```

```
SLAP LACBR
ID SLAP LACBR STANDARD; PRT; 465 AA.
AC Q05044;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE S-layer protein precursor (Surface layer protein).
OS Lactobacillus brevis.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1560;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 8287 / DSM 20556;
RX MEDLINE=93054358; PubMed=1429463;
RA Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
RT "S-layer protein gene of Lactobacillus brevis: Cloning by polymerase
RT chain reaction and determination of the nucleotide sequence.";
RL J. Bacteriol. 174:7419-7427(1992).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
CC S-layer with tetragonal symmetry.
CC -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z14250; CAA78618.1; --
DR PIR; A47023; A47023.
KW Signal; Glycoprotein; Cell wall; S-layer.
FT SIGNAL 1 30
FT CHAIN 31 465
FT SEQUENCE 465 AA; 48159 MW; 2BE2403392E65A2C CRC64;

Query Match
Best Local Similarity 4.9%; Score 121.5; DB 1; Length 465;
Matches 99; Conservative 48; Mismatches 201; Indels 153; Gaps 20;

QY 13 TTTLSESPEN-GIPASWKTIDA----DGDGNWTTTPPPGGTSFAGHNSAICASSASYIN 67
DB 26 TTAQSAKATAGATGAYSTLTKDAATRNVEATGNTALYTKP--GTV---KGAKVVASKATMAK 80
QY 68 FEGQPNPNVLTPELSLPGNGTLTWVCAQDANYASEHYAVYASSTGNDASNFANALLE 127
DB 81 LASSKKSADYFRAYGVKTTNRGSSVYRYVVTMDKGYGVYG-----GKSDTAFAGGI-- 132
QY 128 EVLTAKTVTAPAIRGTRVQGTWYQKTVQ-----LPAGTKYVAFRHFEGCTDFFWINDDD 182
DB 133 -----KSAETTTKADMPARTTGFYLTDSKNTLTAPKYQYKASK----- 173
QY 183 VEIKANGKRADFTTFFSSSTHGRAPAEWTTIDADGQ--QGWLCSSQQLGWLTAHGTPNV 241
DB 174 VSLYGVAKDTKFTVQAAKTKREGSLYVHVVTATNGSGISGWIYAGK--GFSSTATGTGV 230
QY 242 VASFSWNGMALPNPNYLISKDVTAIKVKKYIYAVNDGFPDGHYAVMISKGTGNTAGDFTV 301
DB 231 LGLGLSTDK-----SVTATNDNSVKIYVRTTD-----GTQVGSNTW 266
QY 302 FEETPNGINKGARFGLSTEADGAKPQSVWIER-TVDLPAGTKYVAFRHYNCSD 353
DB 267 -----TSTDGKAGSKVSDKAADQTALEAYINANKPSGTYTNPNA 308
QY 354 DLNY-----ILLDDIQTWGGSPPTDYT-----YTVY 381
DB 309 DATYGNVTYATVSOAATSKVALKVGSGTPVTALTADANDKVAANDTTANGSSVAGSTVY 368
QY 382 RDGTIKIKEGLTETTFF-----EDGVATGNHEVCYEV-----KYTAGVSPK 421
```

```

Db 369 AAGTKLAQLTDLCEKQGVVTLTAIDTLEDAFTGTTTYSDLGKAYHYTYNKSQA 428
Qy 422 ECNVNTVDVQVFNQVNLGTS 442
Db 429 ASSNAS---TOFG--SNVTGT 444

RESULT 8
ID_GUXB_CELFI STANDARD; PRT; 1090 AA.
AC P50839;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
DE 1,4-beta-cellobiohydrolase B (CBP120).
GN CBHB OR CENE.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RN [2]
RN [3]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC STRAIN=ATCC 484;
RX MEDLINE=96003898; PubMed=7575482;
RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulohydrolase B, a second exo-cellobiohydrolase from the
RT cellulolytic bacterium Cellulomonas fimi.";
RL Biochem. J. 311:67-74 (1995).
RN [2]
RN [3]
RP SEQUENCE OF 54-75.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT D (CenD), a family A beta-1,4-glucanase.";
RL J. Bacteriol. 175:1910-1918 (1993).
RN [3]
RP SEQUENCE OF 54-78.
RX MEDLINE=94197708; PubMed=8147863;
RA Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RA Warren R.A.J., Miller R.C. Jr.;
RT "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RT CenB, a member of a new family of beta-1,4-glucanases.";
RL Biochem. Biophys. Res. Commun. 199:1223-1228 (1994).
CC -!- FUNCTION: Hydrolyze cellobiose to a mixture of cellobiose,
CC cellobiose and cellobiose, with only a trace of glucose. It
CC hydrolyzed cellobiose to cellobiose and cellobiose, and
CC cellobiose to cellobiose, but it did not hydrolyze cellobiose.
CC Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
CC with inversion of anomeric configuration.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellobiose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC domain.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; L38827; AB00822.1; -.
CC PR; S59077; S59077.
CC HSP; P07986; 1EXG.
CC InterPro; IPR001919; Bac_celose-bind.
CC InterPro; IPR008965; Cellul_bind.

```

```

DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR00556; Glyco_hydro_48.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR PRINTS; PR00844; GLHYDLASE48.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR SMART; SM00637; CBD_I1; 1.
DR SMART; SM00600; FN3; 3.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 33
FT PROPEP 34 53
FT CHAIN 54 1090
FT DOMAIN 54 699
FT DOMAIN 54 785
FT DOMAIN 700 785
FT DOMAIN 794 884
FT DOMAIN 891 978
FT DOMAIN 989 1090
FT ACT_SITE 513 513
FT DISULFID 990 1089
FT SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;

Query Match 4.9%; Score 121; DB 1; Length 1090;
Best Local Similarity 22.3%; Pred. No. 1.1;
Matches 106; Conservative 53; Mismatches 197; Indels 120; Gaps 26;

Qy 26 ASWKT---IDADGGNNWTTTPPGG-----TSPAGHNSAICASSASINTEGPQNP 75
Db 515 ASWKPSELKWTGKPTWNAAPTGNPLTVETSY-GQDVGAADTARALLFYAAKSGD 573
Qy 76 NYLVTPELSPNGGTLTFWVCAODANVASEHYAVASSTGNDASNFANALLEEVLTAKT 135
Db 574 TASDRKAKALLD---AIWANNQDP-----LGSAVETRGDKRFDPTVAN-----GDG 619
Qy 136 VTAPPAIRGTRVQ-----TWYQKTVQLPAGTKYVAFRHFCTDFFWINLDD 182
Db 620 IYTPSGWTGTPMGNDVKPGVSLDIRSFYKGD---PNWSKVQTFDLDGGAEPQFRVHRFW 676
Qy 183 VEIKANGKRAFTFTFESSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGGTNVV 242
Db 677 AQVAGALADYARLEFDG-----TPTDTAP-----TVPTG---LQAGVWTSTE 719
Qy 243 ASFSWNGMALPNPNYLISKDV-TCAVKYKYVAVNDGFGPDGHYAVMLSKTGTNAGD---F 298
Db 720 ATISWT--ASTDDTRVGYDVYRGATKV-----GTATTSTFDGLTASTAYAY 766
Qy 299 TVVFEETPNNGKGGARFGLSTEA---DGAKPSQWIERVVDLPAGTKYVAFRHYNCSDL 355
Db 767 TVRAFDAAGNVSAALTVTKATPSDTPAP-----SVPAITS-----SSSTA 810
Qy 356 NYILLDDIQFTMGSGPTPTDYT-----YTVVVDGTKIKEGLTETFEEDGV-ATGNHE 407
Db 811 NSV-----TIGWSAS-TDNAGSGLAGYDVYRGATRVQAQ-TTALTFTDTGTASTAYE 861
Qy 408 YCDEVKYTAG--VSPKECVNVTV-----DPVQFNPNQNLGSAVGO-KVTLKWD 454
Db 862 YTVARADVAGNVSAAPSTAVSTVTKSDTPTTTPAPAGLAAMTVETSVALTWNA 917

RESULT 9
CSAA_BACUD
ID_CSAA_BACUD STANDARD; PRT; 1385 AA.
AC Q45760;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry5Aa (Insecticidal delta-endotoxin
DE CryVA(a)) (Crystalline entomocidal protoxin) (152 kDa crystal protein).
GN CRY5AA OR CRYVA(A) OR CRYVA.
OS Bacillus thuringiensis (subsp. darmstadtensis).

```

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 RN NCBI_TaxID=132264;
 RP [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=NRRL B-18243 / PS17;
 RA Sick A.J., Schwab G.E., Payne J.M.;
 RT "Genes encoding nematode-active toxins cloned from *Bacillus*
 RT *thuringiensis* isolate PS17";
 RL Patent number US5281530, 25-JAN-1994.
 CC -|- FUNCTION: ENDOPOXIN WITH NEMATOCIDAL ACTIVITY.
 CC -|- DEVELOPMENTAL STAGE: The crystal protein is produced during
 CC sporulation and is accumulated both as an inclusion and as part of
 CC the spore coat.
 CC -|- MISCELLANEOUS: Toxic segment of the protein is located in the N-
 CC terminus.
 CC -|- SIMILARITY: Belongs to the delta endotoxin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L07025; AAA67694.1; -
 CC PIR; T18213; T18213.
 CC InterPro; IPR005638; endotoxin C.
 CC InterPro; IPR005639; endotoxin N.
 CC InterPro; IPR008979; Gal bind like.
 CC Pfam; PF03944; endotoxin_C; 1.
 CC Pfam; PF03945; endotoxin_N; 1.
 CC Toxin; Sporulation.
 KW Toxin; Sporulation.
 SQ SEQUENCE 1385 AA; 152439 MW; E29FF11FC799DE95 CRC64;
 4.8%; Score 118.5; DB 1; Length 1385;
 Best Local Similarity 20.1%; Pred.No.2.2; Indels 219; Gaps 30;
 Matches 120; Conservative 66; Mismatches 192;
 QY 5 PNPFPNGTTLSEFNGI---PASWKTIDADGDNWNTTPPG---GTSF-----51
 DB 413 PAVDPQAGPNYVSDSNPIQINMTWKTTPQAGS---WNTNLMRGSVSLFLORDGT 470
 QY 52 -----AGNSAICASSA---SYNFEQPNPDNY-----LVPT-ELSLN---87
 DB 471 RLSAGMGGGFADTIYSLPETHYLSYLYGTPYQTSNYSGHVGLVGVSTPQEALENIIG 530
 QY 88 -----GGTLT-FWCAQDA-----NVASEHYAVYAS 112
 DB 531 QPDEQGVSTMGPFPEKASVGGTVVKEWLNANAMKLSFGQSIGIPITNVTGEYQICR 590
 QY 113 STGNDSANF-----ANALLEEVLTAKTVVTAPEAIRTRVQG-----TWYQ 153
 DB 591 YASNDNTNFFNVDTGANGPIFQINFASV-----DNTVGQANGVYVKSIAITDN 644
 QY 154 KTVQLPAGTKYAFRFGCTDFWNLDDVEIKANGKADFTETTFSSHTGERAPAWTTI 213
 DB 645 STEIPAKTINVHLTNQGSDDVF---LDRIEFIPFSLPIYHGSYNTSS-GADDVLW---697
 QY 214 DADGDQGWLCISSLGQLGMLTAGHTGNVVASFSWNGMALNPDNYLISK-----DVTGA 266
 DB 698 -----SSSNVNY-----DIIVGQANSSIASMHLNKGKVIKTIDIPGH 739
 QY 267 TKVKY-YVAVDGFPGDHYAVMISKGTNAGDFTVVFEETPNKGGARFGLSTADGA 325
 DB 740 SETTFATFPVEGF---NEVRILAGLPEVSGNITVQSNPPQPSNNGGDDGGGGG 796
 QY 326 KQSQWIERTVLDPAGTKVAFRHNSDLNIIYLLDDIQFTMGSGSTPT-----374
 DB 797 Q-----YV-----FSLSGSDHTTYHGLEKLEIGH 820
 QY 375 ---DYTYT-----VYRGTGKIKGLTETTF-----EED-----GVATGNHE 407
 DB 407

Db 821 VQGNYYTGTPTVLILNAYRNNTVSSIPVSPDITITQTEADSLLELQPRYGFATVNGT 880
 QY 408 YCVE---VKYTAGVS-PKECVNVT--VDPQVQNPQNLGTSVAVG---QKVTLKWDA 454
 DB 881 ATVKSPNVYDRSFKLPIDLQNTTQNALFASGTQNMIAHNVSDHDIIEVVLKDA 937

RESULT 10
 GUXA CELFI
 ID GUXA CELFI STANDARD; PRT; 872 AA.
 AC P50401; 1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Exoglucanase A precursor (EC 3.2.1.91) (Exocellobiohydrolase A)
 DE (1,4-beta-cellobiohydrolase A) (CBP95).
 GN CBHA.
 OS Cellulomonas fimi.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 484;
 RX MEDLINE=94344030; PubMed=8065260;
 RA Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
 RA Miller R.C. Jr.;
 RT "Cellobiohydrolase A (Cbha) from the cellulolytic bacterium
 RT Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to
 RT Trichoderma reesei CBH II";
 RL Mol. Microbiol. 12:413-422 (1994).
 RN [2]
 RP SEQUENCE OF 41-58.
 RX MEDLINE=93209933; PubMed=8458833;
 RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.,
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 RT D (Cend), a family A beta-1,4-glucanase";
 RL J. Bacteriol. 175:1910-1918 (1993).
 CC -|- FUNCTION: This enzyme hydrolyzes 1,4-beta-D-glucosidic linkages of
 CC cellulose. Weak activity against carboxymethylcellulose, bacterial
 CC microcrystalline cellulose and barley beta-glucan. Has also weak
 CC endoglucanase activity. Hydrolyzes glucosidic bonds with inversion
 CC of anomeric configuration.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -|- SIMILARITY: Contains 3 fibronectin type III domains.
 CC -|- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
 CC domain.
 CC -|- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl
 CC hydrolases).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L25809; AAC36898.1; -
 CC PIR; S49541; S49541.
 CC HSSP; P07986; 1EXG.
 CC InterPro; IPR001919; Bac celose-bind.
 CC InterPro; IPR008965; Cellul bind.
 CC InterPro; IPR008957; FN III-like.
 CC InterPro; IPR003961; FN III.
 CC InterPro; IPR003962; FNIII subd.
 CC InterPro; IPR001524; Glyco_hydro_6.
 CC Pfam; PF00553; CBM 2; 1.
 CC Pfam; PF00041; fn3; 3.
 CC Pfam; PF01341; Glyco_hydro_6; 1.
 CC PRINTS; PR00014; FNTYPEIII.

PRINTS; PR00733; GLYDRIASE6.
DR ProDom; PD003733; Glyco_hydro_6; 2.
DR SMART; SM00637; CBD II; 1.
DR SMART; SM00060; FN3; 3.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 40
FT CHAIN 41 872 EXOGLUCANASE A.
FT DOMAIN 41 477 CATALYTIC
FT DOMAIN 478 563 FIBRONECTIN TYPE-III 1.
FT DOMAIN 573 664 FIBRONECTIN TYPE-III 2.
FT DOMAIN 673 768 FIBRONECTIN TYPE-III 3.
FT DOMAIN 769 872 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT SITE 188 188 PROTON DONOR (BY SIMILARITY).
FT ACT SITE 410 410 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 140 202 BY SIMILARITY.
FT DISULFID 374 428 BY SIMILARITY.
FT DISULFID 770 869 BY SIMILARITY.
SQ SEQUENCE 872 AA; 89300 MW; 7883B407F995533B CRC64;
Query Match 4.6%; Score 114; DB 1; Length 872;
Best Local Similarity 24.3%; Pred. No. 2.4; Indels 112; Gaps 19;
Matches 92; Conservative 29; Mismatches 146; Indels 112; Gaps 19;
Qy 12 GTTILSEFENGIPASWKTIDADGGNNWTTTPPGGTSFAGHNSAICASSASYINFE 71
Db 491 GTTTAT-----SVPLSW---TASTDNVAVTGVDYVGTTLVGTAA----- 528
Qy 72 QNPNDYLTPELSPNGTLTFWCAQDA--NYASEHAVVASSTGNDASFNALLEEV 129
Db 529 ---TSYTVT---GLTPATAYSTFVRKADAGNVSAASAAAATQSGVTDTTAPSV 582
Qy 130 LTAKTVVTAPE-----BAIRGTRVOGTW---YOKTVQLPAGTKYV 165
Db 583 LTAGTTTTTTPLSWTASTDNAGSGVAGYEVLRGTTTGGTTATSYVT-GLRATY- 640
Qy 166 AFRHFGCTDFFWINLDDVEIKANGKRADFTTFE-----SSTHGEA 206
Db 641 -----SFSVRKADVAGNTSAASAASVATQTGTVDVTPAPSVPTGLTAGTTTSSV 691
Qy 207 PAEWTTIADGGQWCLSSQGLWLTAGGTNV--VASFSWNGMALNPDN-----YLIS 260
Db 692 PLTWTF---ASTDNAG---GSGVAGYEVNGTRVATVTSTSYVTGLAADTAYSTVKA 744
Qy 261 KDVTG-----ATKVKYVAVNDGPGDHVAVMLSKGTNAGDFTVVFEETPGINK-G 312
Db 745 KDVAGNVSAASAASVARTQAATSGG-----CTVKYSASSWNTG-FTGTVEVKNNGTAALN 798
Qy 313 GARFGLSTEADGAKPQSVW 331
Db 799 GWTLGFSP-ADGQKVSQGW 816

RESULT 11

YEEU_ECO57
ID YEEU_ECO57 STANDARD; PRT; 2660 AA.
AC Q8X8V7; Q8X2B9; Q8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein YeeU.
GN Z3135 OR ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Enterobacteriaceae; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.,
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RJMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.,
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
frameshift in position 1315.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005423; AAG57041.1; -.
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR000601; PKD.
DR Pfam; PF02369; Big_1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00089; PKD_8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834 BIG-1 1.
FT DOMAIN 840 929 BIG-1 2.
FT DOMAIN 931 1033 BIG-1 3.
FT DOMAIN 1042 1132 BIG-1 4.
FT DOMAIN 1134 1236 BIG-1 5.
FT DOMAIN 1245 1335 BIG-1 6.
FT DOMAIN 1337 1439 BIG-1 7.
FT DOMAIN 1448 1539 BIG-1 8.
FT DOMAIN 1548 1652 BIG-1 9.
FT DOMAIN 1653 1750 BIG-1 10.
FT DOMAIN 1751 1855 BIG-1 11.
FT DOMAIN 1856 1957 BIG-1 12.
FT DOMAIN 1963 2056 BIG-1 13.
FT DOMAIN 2065 2156 BIG-1 14.
FT DOMAIN 2157 2252 BIG-1 15.
FT DOMAIN 2254 2355 BIG-1 16.
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 4.6%; Score 113; DB 1; Length 2660;
Best Local Similarity 22.3%; Pred. No. 12;
Matches 104; Conservative 58; Mismatches 199; Indels 106; Gaps 24;
Qy 13 TTTILSEFE---NGIPASWKTIDADGGNNWTTTPPGGTSFAGHNSAICASSASYINFE 69
Db 1470 TATVKDQFDNEVNLPTVFSTA-----SSGLTLPGESNTNESGLAQTLGVAF----- 1519
Qy 70 GPQPNPNYLTPELSPNGG-----TLTF---WYCAQDANYASEHYAVYASSTGNDASNF 121
Db 1520 GEQ-----TWTASLANNGASDNKTVHFIGDTAAAKIILTTPVDSIAGTPQNSSGS- 1571

QY 122 ANALLEEVLTAKTVVTAPEAIRGTRVOGTWYQKTVOLPAGTKYVAFRHFCTDFFWINLD 181
 Db 1572 -----VITATVVDNNGFVKGVNFTSNAATAEMTNGQAVTNEQGKATVYTNTRS 1624
 QY 182 DVEIKANGKRADETFEFGSTHGEAPAEWTTIDADGGQ-GWCLSGQLGWLTAHGGTN 240
 Db 1625 SIE--SGARPD--TVEASLENGSSTLSTISINADASTAHUTLLIQAUFDTVSAGDTTN 1678
 QY 241 V---VASFSWNG-----MALNPDNLYISKDVGTGATKVKYVAVNDGFGPDHYAVMISK 230
 Db 1679 LYIEVDNNGVPPQEVTLVSP-----SEGVTPSNNAIY-----TTNHDGNEYA---SF 1726
 QY 291 TGTNAGDFTVV-----FEETPNGI-KYGARFGLSTADGAKPQSVMIERTVLDLPAG 341
 Db 1727 TATKAGYQVYATLENGSDSQOVTVPVNVANAEISLAASKDPV-----IANNNDLTLT 1780
 QY 342 TKVAFRHYNC--SDNLVILLDDIQ--FTMGSSPTPTDYTYVYDGTKI---KEGLTE 393
 Db 1781 TATVADTEGNAIANSEVTFPLDPEDVRANFTLG-----DGGKVVTDTEGKAK 1826
 QY 394 TTFEEDGVATGNHEYCEVVKYTAGVSPKCCVNVTVDPVQFNPVQNL 440
 Db 1827 VTLLK--GTRAGAH--TWTASMGAGKSEQLVNVFIADULTIAQVNLNVT 1869
 RESULT 12
 BGAL_BACME STANDARD; PRT; 1034 AA.
 AC 052847;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Beta-galactosidase (EC 3.2.1.23) (Lactase).
 GN BGAM.
 OS Bacillus megaterium.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1404;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 319;
 RA Strey J.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
 CC galactose residues in beta-D-galactosides.
 CC -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ000733; CAA04267.1; -
 CC EIR; T30574; T30574.
 CC HSSP; P00722; IBGL.
 DR InterPro; IPR008979; Gal_bind_like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_2lg.
 DR InterPro; IPR006104; Glyco_hydro_258.
 DR InterPro; IPR006103; Glyco_hydro_2TIM.
 DR InterPro; IPR004200; Glyco_hydro_42C.
 DR InterPro; IPR004199; Glyco_hydro_42N.
 DR Pfam; PF02930; Bgal_small_C; 1.
 DR Pfam; PF02929; Bgal_small_N; 1.
 DR Pfam; PF00703; Glyco_hydro_2; 1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDRASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 KW Hydrolase; Glycosidase.

FT ACT_SITE 481 481 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 547 547 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 1034 AA; 118673 MW; 38644C9A649415E9 CRC64;
 Query Match 4.5%; Score 112.5; DB 1; Length 1034;
 Best Local Similarity 19.9%; Pred. No. 3.9;
 Matches 114; Conservative 70; Mismatches 169; Indels 221; Gaps 35;
 QY 12 GT-TLSEFENGIPAS--WK--TIDA-----DGDGN-----WTTTPPGGTSAGHN 55
 Db 432 GWTYLQEGQKAVPGSKPEWKNVLDRCRSMYERDKNHPSIIWSL-----GNESEGGEN 487
 QY 56 SAIC-----ASSASYINFEPQNPONVLTPELSLPNGGTLTFWVCAODANYASEHYAV 109
 Db 488 FQHMVTFEKEKDSRLVHYHVEGIFHHRDYDASDIES-----TMVYKPADV-----ERYAL 536
 QY 110 -----YASSTGNDASNPAN--ALLEVLTAKTVVTAPAIRGTRVQG-----TWY 152
 Db 537 MNPKEPYILCEYSHAMGNSCGNLYKYWELFDQY-----PILQGGFIWDWK 581
 QY 153 QKTVQLPA--GTKYVAF--RHFGCT-----DFWLNDDVEIKANGKRADETFETESSTHGE 205
 Db 582 DQALQATAEDGTSYLAEGDGFDTPDNGFCNGGLIFADGTAGPKIAEVKKCIQ----- 635
 QY 206 APAEWTTIDADGGQGWCLSGQLGWLTAHGGTNVVA--SFSW-----NG-----MA 251
 Db 636 -PVKMTAVDP-----AKGKPAVQNKHLFTNLNAYDFVMTVEKNGELVEKHASLNLN 684
 QY 252 LMPD-----NYLLSKDVTGAT---KVKY----- 271
 Db 685 VAPDGDDELTLSPLYEQENETDEFVLTLSLRSLKDTAWASAGYEVAYEQFVLPAKAAMP 744
 QY 272 -YYAVNDGFPGDHYAVMISKGTGNAGDFTWVPEETPNGINKGARGFLSTE-----ADGA 325
 Db 745 SVKAAHPALTVDONEQTLTVGIN---FTAIFDK---RKQFTSYNYERTELLASGF 795
 QY 326 KPQSVMIERTVDLPAGTKYVAFRHYNC-----HERCQWRQASLEQHVKKVTVQPDVFVII-SVELAL 848
 Db 796 RP-NFW-RAVTDNDLGNKL---HERCQWRQASLEQHVKKVTVQPDVFVII-SVELAL 848
 QY 368 GGSPTPTDYTYVYRDG-TKIIEGL---TETTFE----- 397
 Db 849 DNSLASCYVITYLYNDGEMKIQSLAPSETPEIPEIGMLFTMNAAFDSLTYWGRPHEN 908
 QY 398 ----EDGVATGNHEYCEVVKYTAGVSPKCCVNV 427
 Db 909 YWDRKTGAKLALHKGSKVEQVTPYLRPQECGNKT 942
 RESULT 13
 ALSI_CANAL STANDARD; PRT; 1260 AA.
 AC P46590;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Agglutinin-like protein 1 precursor.
 GN ALSI.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 11651 / B792;
 RX MEDLINE=95272392; PubMed=7752895;
 RA Hoyer L.L., Scherer S., Shatzman A.R., Iivi G.P.;
 RT "Candida albicans ALSI: domains related to a Saccharomyces cerevisiae
 RT sexual agglutinin separated by a repeating motif";
 RL Mol. Microbiol. 15:39-54(1995).
 CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
 CC -!- PTM: N-glycosylated and O-glycosylated (potential).
 CC -!- SIMILARITY: TO YEAST SAG1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L25902; AAC41649.2; -;
 DR InterPro: IPR008440; Candida ALS.
 DR Pfam: PF05792; Candida ALS; 1.
 KW Cell adhesion; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 1260
 FT DOMAIN 433 792
 FT REPEAT 433 468
 FT REPEAT 469 504
 FT REPEAT 505 540
 FT REPEAT 541 576
 FT REPEAT 577 612
 FT REPEAT 613 648
 FT REPEAT 649 684
 FT REPEAT 685 720
 FT REPEAT 721 756
 FT REPEAT 757 792
 FT DOMAIN 983 1152
 FT REPEAT 983 1043
 FT REPEAT 1092 1152
 FT DOMAIN 399 404
 FT DOMAIN 408 418
 FT DOMAIN 450 455
 FT DOMAIN 486 491
 FT DOMAIN 522 527
 FT DOMAIN 558 563
 FT DOMAIN 594 599
 FT DOMAIN 630 635
 FT DOMAIN 666 671
 FT DOMAIN 702 707
 FT DOMAIN 738 743
 FT DOMAIN 774 779
 FT CARBOHYD 471 471
 FT CARBOHYD 579 579
 FT CARBOHYD 615 615
 FT CARBOHYD 687 687
 FT CARBOHYD 723 723
 FT CARBOHYD 820 820
 FT CARBOHYD 886 886
 FT CARBOHYD 918 918
 FT CARBOHYD 973 973
 FT CARBOHYD 1045 1045
 FT CARBOHYD 1068 1068
 FT SEQUENCE 1260 AA; 132641 MW; 763D1063A2354C24 CRC64;
 Query Match 4.5%; Score 112.5; DB 1; Length 1260;
 Best Local Similarity 19.5%; Pred. No. 5;
 Matches 68; Conservative 30; Mismatches 77; Indels 173; Gaps 16;
 QY 5 PNPENPCTTTL-SSEFENGIPASWKTDADGNN-----ATTVTAPGGDTVIIEPPNHTVTTEYSQSPAT 486
 Db PNPENPCTTTL-SSEFENGIPASWKTDADGNN-----ATTVTAPGGDTVIIEPPNHTVTTEYSQSPAT 486
 QY 41 -TTTPPGGTSFAGHNGAICASSASYINFEQPQP-----DNYLVTPELSLNGG 89
 Db TTTVTAPGGT-----DSVI-----IREPPNPTVTTEYSQSPATTTVTAPGG 532
 QY 90 TLTFW-----CAQDANYASEHYAVYASTGNDASNFANALLEVLTKTVTVAPEAIR 143
 Db TDSVLIIEPPNPTVTTEYSQSYA-----TTTVTAP----- 565
 QY 144 GTRVGQTYQKTVQLPAGTKVAER---HFGCTDFFWLNDDVEIKANGKREADFTTFE 199
 DB GTRVGQTYQKTVQLPAGTKVAER---HFGCTDFFWLNDDVEIKANGKREADFTTFE 199

Db 566 -----PGTDSVLIIEPPNHTVTTEY-----SQSYA 593
 QY 200 SSTHGEAPAEWTTIDADGGQWLCLSSGQLGWLTAHGTTNVASFWSNGMALPNVLI 259
 Db TTTVTAPGGT-----GGTDTV-----IIRPPNHTV 618
 QY 260 SKDVTGATKVKYVAVNDGFGDHYAVMISKGTNAGDFTVFEETPN 307
 Db TTTTEY-----SQSFATTTVTGPPSGDTVIIEPPN 651

RESULT 14

ADP1 MYCGA
 ID ADP1 MYCGA STANDARD; PRT; 1122 AA.
 AC Q49379; Q49437; Q53351;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
 GN GAPA OR MGCI.
 OS Mycoplasma gallisepticum.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S6;
 RX MEDLINE=96201559; PubMed=8613358;
 RA Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
 RT "Cloning and characterization of a putative cytoadhesin gene (mgcl)
 RL from Mycoplasma gallisepticum.";
 RL Infect. Immun. 64:1541-1547(1996).
 [2]
 RP SEQUENCE OF 1-12 FROM N.A.
 RC STRAIN=S6;
 RX Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 378-570 FROM N.A.
 RC STRAIN=S6;
 RX MEDLINE=93371270; PubMed=8363503;
 RA Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
 RT "Identification of the putative cytoadhesin gene of Mycoplasma
 RL gallisepticum and its use as a DNA probe.";
 RL Avian Dis. 37:380-388(1993).
 [4]
 RP SEQUENCE OF 159-1122 FROM N.A.
 RC STRAIN=S6;
 RA Goh M.S., Geary S.J.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Could be involved in cytoadherence.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
 CC -!- SIMILARITY: Belongs to the adhesin P1 family.
 CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
 CC frameshift in position 159 to 213.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U34842; AAC02987.1; -;
 EMBL: U44804; AAC83385.1; ALT_FRAME.
 PIR: T18346; T18346.
 KW Cytoadherence; Signal; Transmembrane.
 FT SIGNAL 1 30
 FT CHAIN 31 1122
 FT TRANSMEM 997 1021
 FT DOMAIN 1002 1009
 FT CONFLICT 313 314
 FT CONFLICT 382 382
 Y -> F (IN REF. 3).


```
FT CONFLICT 501 501 I -> T (IN REF. 3).
FT CONFLICT 568 568 R -> G (IN REF. 3).
FT CONFLICT 570 570 T -> A (IN REF. 3).
FT CONFLICT 693 695 DIL -> V (IN REF. 4).
SQ SEQUENCE 1122 AA; 121305 MW; 155C34DA2D6C65 CRC64;

Query Match
Best Local Similarity 19.0%; Pred. No. 5.9;
Matches 94; Conservative 77; Mismatches 193; Indels 131; Gaps 26;

QY 42 TTPPGGTSPAGHNSAICASSAYINFEQPQEDNVLVTPPELSPNGGTLTFWVCAQDAN 101
Db 437 TBAEPQNTVGPYGLLSA---ISFATR-----NGLALAPALGQDVG 479
QY 102 YASEHYA-----VYASST---GNDASNFAN-----ALLBEVLTKTV 135
Db 480 Y---HFVPRLAGVGVSPRGANGNIFLGSALTWTGTNGNFDLTKWHSPIADPTFT 536
QY 136 VTAPPAIRTRVQGTWYQKTVQLP--AGKYVAFRHFCTDFFWIMLDDEVEIK---ANGK 190
Db 537 VNSSGVLQNS---GSQQSTSTPMNSNGNESIPYRWNTSYDYNVSFAALISKPAAGNTK 593
QY 191 RAD--FTETFESSGTHGEAPEATIID-----ADGGQGWMLCLSSGQLGWLTAHGTTVV 242
Db 594 QVESLETTALKDLTNSLNFKFTQENRIFFSYAMLDGROWSLGTRKDXLWLT-----TNTI 649
QY 243 ASFSMN-----GMLNPNLYLIS-----KDV-----TGATKVKYXYAV 275
Db 650 NNFTYNTQQLASTAAGANANPNILNALTAKGDFDRDIGNVDILYNNNTNKFYYIQV 709
QY 276 ND-----GFPDGHVAMISKTGNAGDTVVVPEETPNINK- GGAARGLSTEADG 324
Db 710 GGAITTWPEVQVNYKTSANITYNLRTDFGSTTPATQDANTVSSKINGAYLSSTGDOQG 769
QY 325 AKPQSVWERTVDLPAGTKYV--AFRHYNCSDLNVLDDIOFTMGSPPTDYTYTVYR 382
Db 770 WTNGSIYVKAESTPSSQGYTWQDFKGLTTASNAVISN---WTKAG-----YSIRP 818
QY 383 DGTIKIEGTTETTFEDBGVATGN-----HEYCVF-----VKYTAGVSPKECVNVTVDPVQ 432
Db 819 DDDTV-FSVSKIPFEKEITAAVNVRLSDSYVQLNGETSVNTVARVSPDSSA-LTLNPKR 876
QY 433 F-NPQVNLGSAVGQ 446
Db 877 ITNPLMN-RDNVIGQ 890

RESULT 15
PME ASPAC
ID PME ASPAC STANDARD; PRT; 331 AA.
AC Q12535;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pectinesterase precursor (EC 3.1.1.11) (Pectin methylesterase) (PE).
GN PME1.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]__TaxID=5053;
RP SEQUENCE FROM N.A.
RC STRAIN=KSM 510;
RX MEDLINE=97079238; PubMed=8920970;
RA Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA Dorrreich K., Dalboege H., Kauppinen S.;
RT "Pectin methyl esterase from Aspergillus aculeatus: expression
RT cloning in yeast and characterization of the recombinant enzyme.";
RL Biochem. J. 319:705-712(1996).
CC -!- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC -!- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC -!- SIMILARITY: Belongs to the pectinesterase family.
-----

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U49378; AAB42153.1; -
DR InterPro; IPR000070; Pectinesterase.
DR Pfam; PF01095; Pectinesterase_1;
DR PROSITE; PS00800; PECTINESTERASE_1; 1.
DR PROSITE; PS00803; PECTINESTERASE_2; 1.
KW Hydrolase; Aspartyl esterase; Cell wall; Signal.
FT SIGNAL 1 17
FT CHAIN 18 331 BY SIMILARITY.
FT ACT_SITE 162 162 BY SIMILARITY.
FT ACT_SITE 183 183 BY SIMILARITY.
SQ SEQUENCE 331 AA; 35681 MW; 1F1C81BF1E32174F CRC64;

Query Match
Best Local Similarity 4.4%; Score 109.5; DB 1; Length 331;
Matches 69; Conservative 33; Mismatches 99; Indels 105; Gaps 15;

QY 108 AVYASSTGNDASNFANALLEVLTKTVVTAPEAIRGTRVQGTWYQKTVQLPAGT-KYVA 166
Db 27 AIVVAKSGDYTTIGDAI--DALSTSTDTQTIFIE---EGT-YDEQVYLPAMTKVII 79
QY 167 FRHFGCTDFFWNL---DDVEIKANGKRADETETFE-----SSTHGEAPAE 209
Db 80 YGQTTNDSYADNLVTITHAISYEDAGESDDLTATFRNKAVGSQVYNLIANTCGAQCHQ 139
QY 210 WTTIDADGGQWML-CLSSGQLGWLTAHGTTNVV----- 242
Db 140 ALALSAMADQQGYGNCFTGYQDTLLAQGNOLYINSYIEGAVDFIFGQHARAFQNVDI 199
QY 243 -----ASFWMGNALNPDN--YLISKDVTGA-----TKVKKYVAVNDGFGPDHYAV 286
Db 200 RVVEGPTASITANGRSSSETDTSYVINKSTVAAKEGDDVARGTYL-----GRPWSEYA- 254
QY 287 MISKGTGNAGDTVVVPEET--PNGINKGARGGLSTEADGAKPQSVWERTVDLPAGTKY 344
Db 255 -----RVVFQQTSMNTNVINSLG-----WTWSTSTP-NTSY 284
QY 345 VAFRHY 350
Db 285 VTFGEY 290

Search completed: May 18, 2004, 11:43:35
Job time : 9.5342 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:36:20 ; Search time 33.3368 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-20

Perfect score: 2480

Sequence: 1 GFPNPNPNPNTTTLSSSF.....QNLGSAVGQKVTLLKWDAPN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 2094 | 84.4 | 1723 | 2 P72197 | P72197 porphyromon |
| 2 | 2082 | 84.0 | 1687 | 2 Q9R9B7 | Q9R9B7 porphyromon |
| 3 | 2082 | 84.0 | 1704 | 2 Q51816 | Q51816 porphyromon |
| 4 | 2079 | 83.8 | 1723 | 2 P72194 | P72194 porphyromon |
| 5 | 2071 | 83.5 | 1358 | 2 P96967 | P96967 porphyromon |
| 6 | 2070 | 83.5 | 1706 | 2 Q51839 | Q51839 porphyromon |
| 7 | 2059 | 83.0 | 1706 | 2 Q51838 | Q51838 porphyromon |
| 8 | 2054.5 | 82.8 | 1223 | 2 Q92NB5 | Q92NB5 porphyromon |
| 9 | 2039.5 | 82.2 | 1097 | 2 P72196 | P72196 porphyromon |
| 10 | 2039.5 | 82.2 | 1732 | 2 O07442 | O07442 porphyromon |
| 11 | 2034.5 | 82.0 | 1732 | 2 Q51817 | Q51817 porphyromon |
| 12 | 2020.5 | 81.5 | 1732 | 2 O52050 | O52050 porphyromon |
| 13 | 682 | 27.5 | 925 | 2 Q9F4J0 | Q9F4J0 porphyromon |
| 14 | 561.5 | 22.6 | 312 | 2 Q9KIB3 | Q9KIB3 porphyromon |
| 15 | 233 | 9.4 | 293 | 2 Q9XB09 | Q9XB09 porphyromon |
| 16 | 152.5 | 6.1 | 1742 | 16 Q9P377 | Q9P377 xanthomonas |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 17 | 144 | 5.8 | 3346 | 16 | Q7WN54 | Q7wn54 bordetella |
| 18 | 139.5 | 5.6 | 1341 | 16 | Q8UAU1 | Q8uau1 agrobacteri |
| 19 | 139 | 5.6 | 2215 | 16 | Q7WBN0 | Q7wbn0 bordetella |
| 20 | 135 | 5.4 | 3988 | 17 | Q8TP21 | Q8tp21 methanosarc |
| 21 | 134.5 | 5.4 | 2768 | 16 | Q8E9G6 | Q8e9g6 shewanella |
| 22 | 134 | 5.4 | 1541 | 16 | Q7UQU9 | Q7uqj9 rhodopirell |
| 23 | 134 | 5.4 | 1800 | 2 | Q9L948 | Q9l948 pseudomonas |
| 24 | 134 | 5.4 | 8682 | 16 | Q88RG2 | Q88rg2 pseudomonas |
| 25 | 133.5 | 5.4 | 680 | 2 | Q52644 | Q52644 ruminococu |
| 26 | 133 | 5.4 | 2522 | 16 | Q8EKA6 | Q8eka6 shewanella |
| 27 | 132.5 | 5.3 | 555 | 16 | Q826W1 | Q826w1 streptomyce |
| 28 | 132.5 | 5.3 | 691 | 16 | Q9RZS7 | Q9rzs7 deinococcus |
| 29 | 132.5 | 5.3 | 1115 | 4 | Q9HD43 | Q9hd43 homo sapien |
| 30 | 131.5 | 5.3 | 1256 | 16 | Q8NX96 | Q8nx96 staphylococ |
| 31 | 131.5 | 5.3 | 1357 | 17 | Q8THC8 | Q8thc8 methanosarc |
| 32 | 131.5 | 5.3 | 2468 | 16 | Q9L2M3 | Q9l2m3 pseudomonas |
| 33 | 131 | 5.3 | 2219 | 16 | Q88W19 | Q88w19 lactobacill |
| 34 | 130.5 | 5.3 | 880 | 2 | Q9RMB8 | Q9rmb8 arthrobacte |
| 35 | 130.5 | 5.3 | 955 | 17 | Q8TQ91 | Q8tg91 methanosarc |
| 36 | 130 | 5.2 | 1386 | 17 | Q8TI72 | Q8ti72 methanosarc |
| 37 | 129.5 | 5.2 | 756 | 9 | Q858B6 | Q858b6 enterobacte |
| 38 | 129.5 | 5.2 | 1744 | 16 | Q82YV8 | Q82yv8 enterococcu |
| 39 | 129 | 5.2 | 7716 | 16 | Q7UMZ8 | Q7umz8 rhodopirell |
| 40 | 128.5 | 5.2 | 1248 | 16 | Q99V41 | Q99v41 staphylococ |
| 41 | 128.5 | 5.2 | 1255 | 2 | Q7WTC6 | Q7wtc6 staphylococ |
| 42 | 128.5 | 5.2 | 2656 | 5 | Q9GDU3 | Q9gnu3 paracentrot |
| 43 | 128 | 5.2 | 1672 | 16 | Q8Y366 | Q8y366 ralstonia s |
| 44 | 128 | 5.2 | 4936 | 16 | Q8YKJ3 | Q8ykj3 anabaena sp |
| 45 | 127.5 | 5.1 | 1684 | 2 | Q03658 | Q03658 unidentifie |

ALIGNMENTS

RESULT 1

P72197 P72197 PRELIMINARY; PRT; 1723 AA.
AC P72197;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Lys-gingipain.
GN KGP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RA Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,
RA Prochazka V., Kiefer M.C., Travis J., Barr P.J.;
RT "Molecular cloning and characterization of Porphyromonas gingivalis
Lys-gingipain."
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; U54691; AAA99810.1; -.
DR MEROPS; C25.002; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000977; DNA ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1723 AA; 16831 MW; 4508A7E50197CEBD CRC64;
Query Match 84.4%; Score 2094; DB 2; Length 1723;
Best Local Similarity 86.2%; Pred. No. 6.8e-128;

[illegible]

Db 1350 ECNVNTINTQFNFKVNLKAQPDGGDVWLKWEAPS 1384

RESULT 3

| | | | |
|--------|---|------|----------|
| Q51816 | PRELIMINARY; | PRT; | 1704 AA. |
| ID | Q51816 | | |
| AC | Q51816; | | |
| DT | 01-NOV-1996 (TrEMBLrel. 01, Created) | | |
| DT | 01-NOV-1996 (TrEMBLrel. 01, Last sequence update) | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | |
| DE | Arg-gingipain-1 proteinase. | | |
| DE | Arg-gingipain-1 proteinase. | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales; | | |
| OC | Porphyromonadaceae; Porphyromonas. | | |
| OX | NCBI_Taxid=837; | | |
| OX | [1] | | |
| RN | SEQUENCE FROM N.A. | | |
| RP | MEDLINE=95138080; PubMed=7836351; | | |
| RX | Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C., | | |
| RA | Travis J., Barr P.J.; | | |
| RT | "Molecular cloning and structural characterization of the Arg- | | |
| RT | gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a | | |
| RT | proteinase-adhesin polypeptide." | | |
| RL | J. Biol. Chem. 270:1007-1010(1995). | | |
| DR | EMBL; U15282; AAA69539.1; -. | | |
| DR | PIR; A55426; A55426. | | |
| DR | HSP; P95493; 1CVR. | | |
| DR | MEROPE; C25.001; -. | | |
| DR | GO; GO:0005524; P:ATP binding; IEA. | | |
| DR | GO; GO:0008234; P:cysteine-type peptidase activity; IEA. | | |
| DR | GO; GO:0003910; P:DNA ligase (ATP) activity; IEA. | | |
| DR | GO; GO:0006310; P:DNA recombination; IEA. | | |
| DR | GO; GO:0006281; P:DNA repair; IEA. | | |
| DR | GO; GO:0006260; P:DNA replication; IEA. | | |
| DR | GO; GO:0006508; P:proteolysis and peptidolysis; IEA. | | |
| DR | InterPro; IPR000977; DNA ligase. | | |
| DR | InterPro; IPR007110; Ig-like. | | |
| DR | InterPro; IPR001769; Peptidase C25. | | |

```

DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA LiGASE_A1; 1.
FT CHAIN 228 719 MATURE 50-KDA CYSTEINE PROTEINASE
SQ SEQUENCE 1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;
Query Match
Best Local Similarity 84.0%; Score 2082; DB 2; Length 1704;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;
QY 3 PNPENPNPGTTLTSESFENGIPASWKTIADGDGNNWTTTPPGTSPAGHNSAICA 62
DB 952 PNPENPNPGTTLTSESFENGIPASWKTIADGDGNGKPGNAPG---IAGYNSNGCV 1008
QY 63 ASY-INFEQPNPDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASF 121
DB 1009 ESFGLGGIGVLTDPNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASF 1068
QY 122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFQCTDFFW 181
DB 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFY 1126
QY 182 DVEIKKANGKRAADFTTFESSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGH 241
DB 1127 EVEIKKANGKRAADFTTFESSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGH 1186
QY 242 VASFSWNGMALPNPNLVLSKDVGTATKVKYKYAVNDGPPGDHYAVMISKTGNAAGFT 301
DB 1187 VASFSWNGMALPNPNLVLSKDVGTATKVKYKYAVNDGPPGDHYAVMISKTGNAAGFT 1246
QY 302 FEETPNKNGKGARFGLSTADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLN 361
DB 1247 FEETPNKNGKGARFGLSTADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLN 1306
QY 362 DIQFTMGSSPTDITYTVYRDTGKIKEGLTTFTEEDGATGNHEYCVVEKYTAGVSP 421
DB 1307 DIQFTMGSSPTDITYTVYRDTGKIKEGLTTFTEEDGATGNHEYCVVEKYTAGVSP 1365
QY 422 ECNVNVTVPQPNVQNLTGSAVGQKVLKWDAPN 456
DB 1367 ECNVNVTINQPNPVKNLKAQPDGGDVVLKWEAPS 1401
RESULT 4
P72194
ID P72194 PRELIMINARY; PRT; 1723 AA.
AC P72194
DT 01-FEB-1997 (TREMELrel. 02, Created)
DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Lys-gingipain.
GN KGP.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97044756; PubMed=8989827;
RA Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
RT "Cloning and sequencing of the gene encoding a novel lysine-specific
RT cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
RT structural relationship with the arginine-specific cysteine proteinase
RT (Arg-gingipain).";
RL J. Biochem 120:398-406 (1996).
DR EMBL; D83258; BAAL1870.1; -.
DR MEROPS; C25.002; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.

DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR005977; DNA_ligase.
DR InterPro; IPR001769; Peptidase_C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
DR PROSITE; PS00697; DNA LiGASE_A1; 1.
SQ SEQUENCE 1723 AA; 187261 MW; 5628963D251493EB CRC64;
Query Match
Best Local Similarity 83.8%; Score 2079; DB 2; Length 1723;
Matches 392; Conservative 17; Mismatches 42; Indels 6; Gaps 4;
QY 1 GTPENPNPGTTLTSESFENGIPASWKTIADGDGNNWTTTPPGTSPAGHNSAICA 60
DB 969 GTPENPNPGTTLTSESFENGIPASWKTIADGDGNGKPGNAPG---IAGYNSNGCV 1025
QY 61 SSASY-INFEQPNPDNVLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
DB 1026 YSEFGLGGIGVLTDPNVLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085
QY 120 NFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHFQCTDFF 179
DB 1086 NFNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFY 1143
QY 180 LDDVEIKKANGKRAADFTTFESSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGH 239
DB 1144 LDEVEIKKANGKRAADFTTFESSSTHGEAPAEWTTIDADGGQGWCLSSGQGLWLTAGH 1203
QY 240 NVASFSWNGMALPNPNLVLSKDVGTATKVKYKYAVNDGPPGDHYAVMISKTGNAAGFT 299
DB 1204 NVASFSWNGMALPNPNLVLSKDVGTATKVKYKYAVNDGPPGDHYAVMISKTGNAAGFT 1263
QY 300 VFETPNKNGKGARFGLSTADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLN 359
DB 1264 VFETPNKNGKGARFGLSTADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLN 1323
QY 360 LDDQFTMGSSPTDITYTVYRDTGKIKEGLTTFTEEDGATGNHEYCVVEKYTAGV 419
DB 1324 LDDQFTMGSSPTDITYTVYRDTGKIKEGLTTFTEEDGATGNHEYCVVEKYTAGV 1383
QY 420 PRECVNVTVPQPNVQNLTGSAVGQKVLKWDAPN 456
DB 1384 PKVCNVNVTINQPNPVKNLKAQPDGGDVVLKWEAPS 1420
RESULT 5
P96967
ID P96967 PRELIMINARY; PRT; 1358 AA.
AC P96967
DT 01-MAY-1997 (TREMELrel. 03, Created)
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAGD.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA Han N., Lepine G., Whitlock J., Wojciechowski L., Proguleske-Fox A.;
RT "Cloning, sequencing and characterization of hagD, a member of the
RT Harep multigene family in Porphyromonas gingivalis."
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U68468; AAB49691.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.

```

```

DR GO:0006310; P:DNA recombination; IEA.
DR GO:0006281; P:DNA repair; IEA.
DR GO:0006260; P:DNA replication; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000977; DNA_ligase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001769; Peptidase_C25.
DR InterPro: IPR005536; Peptidase_C25_C.
DR Pfam: PF01364; Peptidase_C25; I.
DR Pfam: PF03785; Peptidase_C25_C; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS00697; DNA_LIGASE_A1; 1.
SQ SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;

Query Match      83.5%; Score 2071; DB 2; Length 1358;
Best Local Similarity 85.6%; Pred. No. 1.6e-126;
Matches 391; Conservative 17; Mismatches 43; Indels 6; Gaps 4;

QY 1 GTPNPNPNPGTTTSSSEFENGIPASWKTIDADGNNWTTTPPGGTSGAGHNSAICA 60
DB 604 GTPNPNPNPGTTTSSSEFENGIPASWKTIDADGNNWTTTPPGGTSGAGHNSAICA 60
QY 61 SSASY-INFEQPNPDNYLTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
DB 661 YSSEFGLGGIGVLTDPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 720
QY 120 NFANALLEEVLTAQVTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFEWIN 179
DB 721 NFTNALLEEITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQCTDFEWIN 778
QY 180 LDDVEIKANGKRAQDTFTFSSSTHGEAPAEWTTIDADGGQWCLSSGQLGWLTAHGGT 239
DB 779 LDEVEIKANGKRAQDTFTFSSSTHGEAPAEWTTIDADGGQWCLSSGQLGWLTAHGGT 838
QY 240 NVVASFSGWNGMALNPNDNYLISKDVTGATKYKYVAVNDGPGDHYAVMISKTGTNAGDFT 299
DB 839 NVVASFSGWNGMALNPNDNYLISKDVTGATKYKYVAVNDGPGDHYAVMISKTGTNAGDFT 898
QY 300 VFEETENGINKGARGFLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNYYL 359
DB 899 VFEETENGINKGARGFLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNYYL 958
QY 360 LDDIQTFTMGSGPTPTDYYTVYRDGTKIKEGLTFTFEEDGATGNHEVCYEVKYTAGVS 419
DB 959 LDDIQTFTMGSGPTPTDYYTVYRDGTKIKEGLTFTFEEDGATGNHEVCYEVKYTAGVS 1018
QY 420 PKECVNTVDPVQFPNPQNLTGSAVGQKVTLKWDAPN 456
DB 1019 PKVCVNTINPTQFPNPVKNLKAQPDGDDVVLKWEAPS 1055

RESULT 6
QY Q51839 PRELIMINARY; PRT; 1706 AA.
AC Q51839; Q51840;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Arginine-specific thiol protease precursor.
GN PRTR.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RX MEDLINE=95160709; PubMed=7857299;
RA Kirsztbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
RA Reynolds E.C.;
RT "Complete nucleotide sequence of a gene prtr of Porphyromonas
RT gingivalis W50 encoding a 132 kDa protein that contains an arginine-
RT specific thiol endopeptidase domain and a haemagglutinin domain.";

```

```

RL Biochem. Biophys. Res. Commun. 207:424-431(1995).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=W50;
RX MEDLINE=96311339; PubMed=8713096;
RA Slakeski N., Cleal S.M., Reynolds E.C.;
RT "Characterization of a Porphyromonas gingivalis gene prtr that encodes
RT an arginine-specific thiol proteinase and multiple adhesins.";
RL Biochem. Biophys. Res. Commun. 224:605-610(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Reynolds E.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Slakeski N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L26341; AAC18876.1; -.
DR HSP; P95493; 1CVR.
DR GO:0005524; F:ATP binding; IEA.
DR GO:0008234; F:cysteine-type peptidase activity; IEA.
DR GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR GO:0006310; P:DNA recombination; IEA.
DR GO:0006281; P:DNA repair; IEA.
DR GO:0006508; P:DNA replication; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000977; DNA_ligase.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR001769; Peptidase_C25.
DR InterPro: IPR005536; Peptidase_C25_C.
DR Pfam: PF01364; Peptidase_C25; I.
DR Pfam: PF03785; Peptidase_C25_C; 1.
DR PROSITE: PS00697; DNA_LIGASE_A1; 1.
KW Protease; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 228 1706 ARGinine-SPECIFIC THIOL PROTEASE.
SQ SEQUENCE 1706 AA; 185626 MW; E8BDF07C9813B844 CRC64;

Query Match      83.5%; Score 2070; DB 2; Length 1706;
Best Local Similarity 85.7%; Pred. No. 2.5e-126;
Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTSSSEFENGIPASWKTIDADGNNWTTTPPGGTSGAGHNSAICASS 62
DB 954 PNPENPNPGTTTSSSEFENGIPASWKTIDADGNNWTTTPPGGTSGAGHNSAICASS 1010
QY 63 ASY-INFEQPNPDNYLTPELSLPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
DB 1011 ESFGLGGIGVLTDPDNYLITPALDLANGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070
QY 122 ANALLEEVLTAQVTVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFEWINLD 181
DB 1071 TNALLEEITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQCTDFEWINLD 1128
QY 182 DVEIKANGKRAQDTFTFSSSTHGEAPAEWTTIDADGGQWCLSSGQLGWLTAHGGTNNV 241
DB 1129 DVEIKANGKRAQDTFTFSSSTHGEAPAEWTTIDADGGQWCLSSGQLGWLTAHGGTNNV 1188
QY 242 VASFSWNGMALNPNDNYLISKDVTGATKYKYVAVNDGPGDHYAVMISKTGTNAGDFTVV 301
DB 1189 VSSFSWNGMALNPNDNYLISKDVTGATKYKYVAVNDGPGDHYAVMISKTGTNAGDFTVV 1248
QY 302 FEETPNKNGKARGFLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNYYL 361
DB 1249 FEETPNKNGKARGFLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNYYL 1308
QY 362 DIQFTMGSGPTPTDYYTVYRDGTKIKEGLTFTFEEDGATGNHEVCYEVKYTAGVSPK 421
DB 1309 DIQFTMGSGPTPTDYYTVYRDGTKIKEGLTFTFEEDGATGNHEVCYEVKYTAGVSPK 1368
QY 422 ECNVNTVDPVQFPNPQNLTGSAVGQKVTLKWDAPN 456

```

| | | | | | | |
|--|--|---|---------------|------|--|--|
| Db | 1369 | KCVNVTNSTQFNPKLKAQPDGSDVVLKWEAPS | 1403 | 1129 | EVEIKANGKRADFTETFSSTHGEATAEWTTIDADGQGWCLSSGQDLTAHGCTNV | 1188 |
| RESULT 7 | | | | | | |
| ID | Q51838 | PRELIMINARY; | PRT; 1706 AA. | | | |
| AC | Q51838; | | | | | |
| DT | 01-NOV-1996 (TremBLrel. 01, Created) | | | | | |
| DT | 01-MAY-1997 (TremBLrel. 03, Last sequence update) | | | | | |
| DT | 01-OCT-2003 (TremBLrel. 25, Last annotation update) | | | | | |
| DE | Protease precursor. | | | | | |
| GN | PRP1. | | | | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | | | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales; | | | | | |
| OC | Porphyromonadaceae; Porphyromonas. | | | | | |
| OX | NCBI_TaxID=837; | | | | | |
| RN | [1] _ | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RC | STRAIN=W50; | | | | | |
| RX | MEDLINE=96071894; PubMed=7591131; | | | | | |
| RA | Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.; | | | | | |
| RT | "Characterization, genetic analysis, and expression of a protease | | | | | |
| RT | antigen (Prp1) of Porphyromonas gingivalis W50."; | | | | | |
| RL | Infect. Immun. 63:4744-4754 (1995). | | | | | |
| RN | [2] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RC | STRAIN=W50; | | | | | |
| RA | Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.; | | | | | |
| RT | "The prp1 and the prp2 arginine-specific protease genes of | | | | | |
| RT | porphyromonas gingivalis W50 produce five biochemically distinct | | | | | |
| RT | enzymes."; | | | | | |
| RL | Mol. Microbiol. 23:0-0 (1997). | | | | | |
| DR | EMBL; X82680; CAA57997.1; - | | | | | |
| DR | HSSP; P95493; 1CVR. | | | | | |
| DR | MEROES; C25.001; - | | | | | |
| DR | GO; GO:000524; P:ATP binding; IEA. | | | | | |
| DR | GO; GO:0008234; P:cysteine-type peptidase activity; IEA. | | | | | |
| DR | GO; GO:0003910; F:DNA ligase (ATP) activity; IEA. | | | | | |
| DR | GO; GO:0006310; P:DNA recombination; IEA. | | | | | |
| DR | GO; GO:0006281; P:DNA repair; IEA. | | | | | |
| DR | GO; GO:0006260; P:DNA replication; IEA. | | | | | |
| DR | GO; GO:0006508; P:proteolysis and peptidolysis; IEA. | | | | | |
| DR | InterPro; IPR000977; DNA ligase. | | | | | |
| DR | InterPro; IPR007110; IG-like. | | | | | |
| DR | InterPro; IPR001769; Peptidase C25. | | | | | |
| DR | InterPro; IPR005536; Peptidase C25_C. | | | | | |
| DR | Pfam; PF01364; Peptidase C25; 1. | | | | | |
| DR | Pfam; PF03785; Peptidase C25 C; 1. | | | | | |
| DR | PROSITE; PS00697; DNA_LIGASE_A1; 1. | | | | | |
| KW | Signal. | | | | | |
| FT | SIGNAL. | | | | | |
| FT | CHAIN | | | | | |
| FT | CHAIN | | | | | |
| FT | CHAIN | | | | | |
| SQ | SEQUENCE | 1706 AA; 185705 MW; 0E56DCD87FDA8CDD | CRC64; | | | |
| Query Match | | | | | | |
| Best Local Similarity | | | | | | |
| Matches 389; Conservative 17; Mismatches 43; Indels 6; Gaps 4; | | | | | | |
| QY | 3 | PNPENPNPNTLTSESFENGIPASWKITIDADGDNNTTTPPPGGTSFAGHNSAICASS | 62 | QY | 1 | GTPNPNPNPNTLTSESFENGIPASWKITIDADGDNNTTTPPPGGTSFAGHNSAIC |
| Db | 954 | PNPENPNPNTLTSESFENGIPASWKITIDADGDNNTTTPPPGGTSFAGHNSAIC | 1010 | Db | 460 | GTPNPNPNPNTLTSESFENGIPASWKITIDADGDNNTTTPPPGGTSFAGHNSAIC |
| QY | 63 | ASY-INFECPNPNDNYLTPPELSLNGGTLTPWCAQDANYASEHYAVASSTGNDASNF | 121 | QY | 60 | ASSASY-INFECPNPNDNYLTPPELSLNGGTLTPWCAQDANYASEHYAVASSTGND |
| Db | 1011 | BSFGLGGIGLVLPDNYLTPPELSLNGGTLTPWCAQDANYASEHYAVASSTGNDASNF | 1070 | Db | 517 | VYSFGLGGIGLVLPDNYLTPPELSLNGGTLTPWCAQDANYASEHYAVASSTGND |
| QY | 122 | ANALLEEVLTAKTVTAPETRGVQGTWQKTQVLPAGTKYVAFRFGCTDFWIMLD | 181 | QY | 119 | SNEANALLEEVLTAKTVTAPETRGVQGTWQKTQVLPAGTKYVAFRFGCTDFWIM |
| Db | 1071 | TNALLEETITAKG-VRSPEARIG-RIOGTWRQKTVDLPAGTKYVAFRFGCTDFWIM | 1128 | Db | 577 | SNFNALLEETITAKG-VRSPEARIG-RIOGTWRQKTVDLPAGTKYVAFRFGCTDFWIM |
| QY | 182 | DVEIKANGKRADFTETFSSTHGEATAEWTTIDADGQGWCLSSGQDLTAHGCTNV | 241 | | | |

QY 60 ASSASY-INFEQPNPDNYLTPVPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGND 118
 Db 1026 VYSEFGLGGIGVLTDPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGND 1085
 QY 119 SNFANALLEVLTAKTIVTAPAIKRTVRQGTWYQKTVOLPAGTKYVAFRHHGCTDFFWI 178
 Db 1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVLDPAGTKYVAFRHHGCTDFFWI 1143
 QY 179 NLDDVEIKANGKRAFTTFEFSSTHGEAPAEWTTTIDADGGGWLCLSSGQDLWLTAGG 238
 Db 1144 DLDEVEIKANGKRAFTTFEFSSTHGEAPAEWTTTIDADGGGWLCLSSGQDLWLTAGG 1203
 QY 239 TNVAFSFWNGMALNPDNYLISKDVTGATKVKYKYYAVNDGFFGDHYAVMISKTGNAGDF 298
 Db 1204 SNVVSFSWNGMALNPDNYLISKDVTGATKVKYKYYAVNDGFFGDHYAVMISKTGNAGDF 1263
 QY 299 TVVFETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHHNCSDLNYI 358
 Db 1264 TVVFETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHHNCSDLNYI 1323
 QY 359 LLDDIOFTMGSGPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHCVEVKYTAGV 418
 Db 1324 LLDDIOFTMGSGPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHCVEVKYTAGV 1383
 QY 419 SPKECVNVTVDVQFNPVQNLTSAGV--GQKVLKWDAP 455
 Db 1384 SPKCKVDVTNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422

RESULT 11
 Q51817
 ID Q51817 PRELIMINARY; PRT; 1732 AA.
 AC Q51817;
 DT 01-NOV-1996 (TremBLrel. 01, Created)
 DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Porphyropain.
 GN PRTP.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W12;
 RX MEDLINE=96213011; PubMed=8631659;
 RA Bartocoy-Gallagher G.A., Han N., Patti J.M., Whitlock J.,
 RA Proguilake-Fox A., Lantz M.S.;
 RT "Analysis of the prtp gene encoding porphyropain, a cysteine proteinase
 of Porphyromonas gingivalis."
 RL J. Bacteriol. 178:2734-2741(1996).
 DR EMBL; U42210; AAB06565.1; --
 DR PIR; T30836; T30836.
 DR MEROPS; C25.002; --
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25_1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1732 AA; 187875 MW; 654271DBEF7BCA54 CRC64;

Query Match 82.0%; Score 2034.5; DB 2; Length 1732;
 Best Local Similarity 84.3%; Pred. No. 5.2e-124;
 Matches 397; Conservative 19; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPCT-TTLSESPENGIPASWKTTIDADGNNWTTTPPGGTSFAGHNSAIC 59
 Db 969 GTPNPNPNPNPCT-TTLSESPENGIPASWKTTIDADGNNWTTTPPGGTSFAGHNSAIC 1025
 QY 60 ASSASY-INFEQPNPDNYLTPVPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGND 118
 Db 1026 VYSEFGLGGIGVLTDPDNYLIITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGND 1085
 QY 119 SNFANALLEVLTAKTIVTAPAIKRTVRQGTWYQKTVOLPAGTKYVAFRHHGCTDFFWI 178
 Db 1086 SNFTNALLEETITAKG-VRSPKAIRG-RIQGTWRQKTVLDPAGTKYVAFRHHGCTDFFWI 1143
 QY 179 NLDDVEIKANGKRAFTTFEFSSTHGEAPAEWTTTIDADGGGWLCLSSGQDLWLTAGG 238
 Db 1144 DLDEVEIKANGKRAFTTFEFSSTHGEAPAEWTTTIDADGGGWLCLSSGQDLWLTAGG 1203
 QY 239 TNVAFSFWNGMALNPDNYLISKDVTGATKVKYKYYAVNDGFFGDHYAVMISKTGNAGDF 298
 Db 1204 SNVVSFSWNGMALNPDNYLISKDVTGATKVKYKYYAVNDGFFGDHYAVMISKTGNAGDF 1263
 QY 299 TVVFETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHHNCSDLNYI 358
 Db 1264 TVVFETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHHNCSDLNYI 1323
 QY 359 LLDDIOFTMGSGPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHCVEVKYTAGV 418
 Db 1324 LLDDIOFTMGSGPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHCVEVKYTAGV 1383
 QY 419 SPKECVNVTVDVQFNPVQNLTSAGV--GQKVLKWDAP 455
 Db 1384 SPKCKVDVTNSTQFNPVQNLTAEQAPNSMDAILKWNAP 1422

RESULT 12
 OS2050
 ID OS2050 PRELIMINARY; PRT; 1732 AA.
 AC OS2050;
 DT 01-JUN-1998 (TremBLrel. 06, Created)
 DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Lysine specific cysteine protease.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W83;
 RX MEDLINE=98298016; PubMed=9632563;
 RA Lewis J.P., Macrina F.L.;
 RT "IS195, an insertion sequence-like element associated with protease
 genes in Porphyromonas gingivalis."
 RL Infect. Immun. 66:3035-3042(1998).
 DR EMBL; AF017059; AAC26523.1; --
 DR MEROPS; C25.002; --
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25_1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1732 AA; 187931 MW; B2337463D5CB5EA5 CRC64;

Query Match 81.5%; Score 2020.5; DB 2; Length 1732;

Best Local Similarity 83.9%; Pred. No. 4.2e-123;
Matches 385; Conservative 19; Mismatches 46; Indels 9; Gaps 6;

QY 1 GTPNPNPNPCT-TTLESFENGIPASWKITDADGDGNNWTTTPPGGTSPAGHNSAIC 59
DB 969 GTPNPNPNPNPCT-TTLESFENGIPASWKITDADGDGNNWTTTPPGGTSPAGHNSAIC 1025

QY 60 ASSASY-INFGSPQNDNLTPELSLPGGTLTFWVCAQDANYASEHYAYASSTGND 118
DB 1026 VYSEFGGGGLVLPDNLITPDLNPGGKLTTFWVCAQDANYASEHYAYASSTGND 1085

QY 119 SNFANALLEEVLTAQVTVTAPEAIRGRVQSTWYQKTVQLPAGTKYVAFRHFGCTDFEWI 178
DB 1086 SNFTNALLEETITAGK-VRSKPAIRG-RIQGTWRQKTVQLPAGTKYVAFRHFGCTDFEWI 1143

QY 179 NLDVEIKANGKRAQDFTTFSSSTHGEAPAEWTTTIDADGGQGMCLSSGQGLMLTAHGG 238
DB 1144 DLDEVEIKANGKRAQDFTTFSSSTHGEAPAEWTTTIDADGGQGMCLSSGQGLMLTAHGG 1203

QY 239 TNVVASPSWNGMALNPDNLYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 298
DB 1204 SNVSSPSWNGMALNPDNLYLSKDVGTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 1263

QY 299 TVVFEETPNGKNGARGLSTEADGAKPQSWIERTVDLPAGTKYVAFRHNGSCLNYI 358
DB 1264 TVVFEETPNGKNGARGLSTEADGAKPQSWIERTVDLPAGTKYVAFRHNGSCLNYI 1323

QY 359 LLDIOFTMGSSPTDYYTVYRDGKIKEGLTETFEEDGATGNHEVCVEVKYTAGV 418
DB 1324 LLDIOFTMGSSPTDYYTVYRDGKIKEGLTETFEEDGATGNHEVCVEVKYTAGV 1383

QY 419 SPKECVNVTVDPVQNPVQNLITGSAV--GQVTLKWDAP 455
DB 1384 SPKECVNVTVDPVQNPVQNLITGSAV--GQVTLKWDAP 1422

RESULT 13
QYF4J0 PRELIMINARY; PRT; 925 AA.
AC QYF4J0;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE Putative outer membrane protein PG57.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT "Identification of vaccine candidates from genomic analysis of
Porphyromonas gingivalis.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007522; AAG24228.1; --
DR InterPro; IPR003961; FN_III.
DR SMART; SM00060; PN3; 1.
SQ SEQUENCE 925 AA; 103632 MW; 5FF2198D6914DAE8 CRC64;

Query Match 27.5%; Score 682; DB 2; Length 925;
Best Local Similarity 30.2%; Pred. No. 4e-36;
Matches 176; Conservative 62; Mismatches 132; Indels 212; Gaps 22;

QY 16 LSESFENGIPASWKITDADGDGNNWTTTPPGG-----TSFAGHNSAICASSASYINF 68

DB 311 LYENFENGPVNGMLVADADGGSFWS-----GHYNAYDAFFPHNGGHSLSASYVPG 363
QY 69 EGQPNPNLYTPELSLPGGTLTFWVCAQDANYASEHYAYASSTGNDASNFANALLEE 128
DB 364 IGPVTPDNLITPKE--GAKRVKYVWSTQDANWAAEHYAVMASTTGTAVGDFV-ILFEE 420

QY 129 VLTAKTVTAPEAIRGRVQSTWYQKTVQLPAGTKYVAFRHFGCTDFEWNLNDV----- 183
DB 421 TMTAKPT-----GAWYERTINLPGTKYIAWRHYNCTDIYFLKLDITVFGT 467

QY 184 -----EIKANG----- 189
DB 468 PASEPEPVTVDFVSLIENKGRKLNWPNPGVEYEDKDDKPLQLAGYNIYANGSLLVHIQ 527

QY 190 -----KRAQDFTET-----FRSSTHGE 205
DB 528 DPTVLEYIDETYSRDKDQVEYCVTAVYNDNIESQVCDKLIYDSQDILYSGFEAGS 587

QY 206 APAEWTTIDADGGQGMCLSSGQGLM-LTAHGTVNVVASFWSNGM--ALNPDNLYLSKD 262
DB 588 IPEGLMLIDADGDNVW-----DYYPWTMYGHDSEKCIASESYLPMIGVLTDPNLYVTPR 642

QY 263 VTGATKVKYVAVNDG--FPGDHYAVMISKTGTNAGDFTVVFETPNGKNGARGLSTE 321
DB 643 LEGAKLVYWSAQDANYASEHYAVMASTTGTAVEDFVLLFEET-----MTAK 690

QY 322 ADGAKPQSWIERTVDLPAGTKYVAFRHNGSCLNYILLDDI----- 363
DB 691 ANGA-----WYERTITLPAQTKYIAWRHYDCTDMFFLLDDITVYRSTETVPEVPTDFV 745

QY 364 -----QFTMGSSPTPD-----YTVTVYRDGT-----KIKEGLTETTF 396
DB 746 SLIENKGRKLNWPNPGVEYEDKDDKPLQLAGYNIYANGSLLVHIQDPTVLEYIDETYS 805

QY 397 EEDGATGNHEVCVEVKYTAGV-SPKEC--VNVTV--DPVQ 432
DB 806 SRDQGV--EMEYCVTAVYNDNIESQVCDKLIYDITSLDNIQ 845

RESULT 14
QYKIB3 PRELIMINARY; PRT; 312 AA.
AC QYKIB3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237555; AAF81413.1; --
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match 22.6%; Score 561.5; DB 2; Length 312;
Best Local Similarity 32.8%; Pred. No. 7e-29;
Matches 136; Conservative 23; Mismatches 73; Indels 183; Gaps 8;

QY 16 LSESFENGIPASWKITDADGDGNNWTTTPPGGTSPAGHNSAICASSASYINFEGQNDP 75

[illegible]

RESULT 15

| | | | | | |
|--------|---|---|--------------------|---------|----------------------------------|
| Q9XBU9 | | PRELIMINARY; | PRT; | 293 AA. | |
| ID | Q9XBU9 | | | | |
| AC | Q9XBUE9; | | | | |
| DT | 01-NOV-1999 | (TrEMBLrel. 12, Created) | | | |
| DT | 01-NOV-1999 | (TrEMBLrel. 12, Last sequence update) | | | |
| DT | 01-OCT-2002 | (TrEMBLrel. 22, Last annotation update) | | | |
| DE | Immunoreactive 32 kDa antigen PG25. | | | | |
| OS | Porphyromonas gingivalis (Bacteroides gingivalis). | | | | |
| OC | Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales; | | | | |
| OC | Porphyromonadaceae; Porphyromonas. | | | | |
| OX | NCBI_TaxID=837; | | | | |
| [1] | | | | | |
| RN | SEQUENCE FROM N.A. | | | | |
| RP | RC STRAIN=WSO; | | | | |
| RA | Roos B., Barr I., Patterson M., Agius C., Rothel L., Margetts M., | | | | |
| RA | Hocking D., Webb E.; | | | | |
| RT | "P. gingivalis polypeptides and nucleic acids.;" | | | | |
| EL | Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| DR | EMBL; AF155351; AAD38410.1; -- | | | | |
| SQ | SEQUENCE 293 AA; 3272 MW; CC03EAC241F7F6F1 CRC64; | | | | |
| | Query Match | 9.4%; | Score 233; | DB 2; | Length 293; |
| | Best Local Similarity | 59.0%; | Fred. No. 1.7e-07; | | |
| | Matches | 46; | Conservative | 7; | Mismatches 23; Indels 2; Gaps 1; |
| Qy | 376 YTYTVYRDGTKKGLTETTFEEVDGVATGNHEYCVVEKYTAGVSKECWNVTVPDVFNP | 435 | | | |
| Dd | 77 YTYTVYRDGVKIAGLTETPTFLDEDVPAGETHYCVEVOYQGVSDDKVCVDVEVK--DFKP | 134 | | | |
| Qy | 436 VONLTGSavgqkvTLKWd | 453 | | | |
| Dd | 135 VTNLGTgASNDvSLDWd | 152 | | | |

Search completed: May 18, 2004, 11:46:39
Job time : 34.3368 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 14.0888 Seconds
(without alignments)
1670.936 Million cell updates/sec

Title: US-08-570-311-20

Perfect score: 2480

Sequence: 1 GTPNPNPNPGTTILSESF.....QNLTGSAVGQKVTWKDAFN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pap.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pap.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pap.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pap.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pap.*
- 6: /cgn2_6/prodata/2/iaa/backfile1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------------|
| 1 | 2480 | 100.0 | 456 | 2 | US-08-570-311-18 |
| 2 | 2480 | 100.0 | 456 | 2 | US-08-570-311-20 |
| 3 | 2480 | 100.0 | 2628 | 2 | US-08-570-311-14 |
| 4 | 2436 | 98.2 | 450 | 2 | US-08-570-311-16 |
| 5 | 2351 | 94.8 | 439 | 2 | US-08-570-311-22 |
| 6 | 2082 | 84.0 | 1477 | 4 | US-09-482-500A-1 |
| 7 | 2082 | 84.0 | 1687 | 2 | US-08-570-311-29 |
| 8 | 2082 | 84.0 | 1704 | 3 | US-08-336-308A-10 |
| 9 | 2082 | 84.0 | 1704 | 3 | US-08-822-324-6 |
| 10 | 2082 | 84.0 | 1704 | 3 | US-09-490-931-10 |
| 11 | 2070 | 83.5 | 1706 | 4 | US-09-066-330-10 |
| 12 | 2066 | 83.3 | 1087 | 2 | US-08-570-311-8 |
| 13 | 2066 | 83.3 | 1087 | 2 | US-08-353-485-8 |
| 14 | 2066 | 83.3 | 1358 | 2 | US-08-570-311-27 |
| 15 | 2039.5 | 82.2 | 1732 | 4 | US-09-066-330-11 |
| 16 | 2034.5 | 82.0 | 1732 | 2 | US-08-570-311-10 |
| 17 | 2034.5 | 82.0 | 1732 | 2 | US-08-353-485-10 |
| 18 | 546.5 | 22.0 | 497 | 2 | US-08-570-311-2 |
| 19 | 546.5 | 22.0 | 497 | 2 | US-08-353-485-2 |
| 20 | 296.5 | 12.0 | 942 | 1 | US-08-141-324-14 |
| 21 | 296.5 | 12.0 | 942 | 1 | US-08-541-902-14 |
| 22 | 236 | 9.5 | 49 | 3 | US-08-822-324-18 |
| 23 | 189 | 7.6 | 46 | 3 | US-08-822-324-9 |
| 24 | 137 | 5.5 | 1833 | 4 | US-08-621-944A-4 |
| 25 | 137 | 5.5 | 1833 | 4 | US-08-945-567D-4 |
| 26 | 137 | 5.5 | 1992 | 4 | US-08-621-944A-3 |
| 27 | 137 | 5.5 | 1992 | 4 | US-08-945-567D-3 |

28 134.5 5.4 2736 4 US-09-252-991A-30227 Sequence 30227, A
29 132.5 5.3 2315 4 US-09-543-681A-5434 Sequence 5434, A
30 131.5 5.3 492 4 US-09-482-500A-2 Sequence 2, Appli
31 131.5 5.3 737 1 US-08-119-361-5 Sequence 5, Appli
32 131.5 5.3 737 3 US-08-336-308A-4 Sequence 4, Appli
33 131.5 5.3 737 3 US-08-822-324-4 Sequence 4, Appli
34 131.5 5.3 737 3 US-09-490-931-4 Sequence 4, Appli
35 131.5 5.3 2123 3 US-08-968-685A-10 Sequence 10, Appli
36 129.5 5.2 1638 4 US-09-071-035-258 Sequence 258, App
37 129.5 5.2 1638 4 US-09-071-035-266 Sequence 262, App
38 129.5 5.2 1747 4 US-09-134-000C-5999 Sequence 5999, Ap
39 129.5 5.2 24 4 US-09-066-330-6 Sequence 6, Appli
40 129 5.2 24 3 US-08-822-324-19 Sequence 19, Appli
41 128 5.1 2411 4 US-09-268-347-36 Sequence 36, Appli
42 126 5.0 669 4 US-09-071-035-264 Sequence 264, App
43 123.5 5.0 2048 4 US-09-268-347-48 Sequence 48, Appli
44 123.5 5.0 509 3 US-08-822-324-8 Sequence 8, Appli
45 121

ALIGNMENTS

RESULT 1
US-08-570-311-18
; Sequence 18, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Gylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 456 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-18

Query Match
Best Local Similarity 100.0%; Score 2480; DB 2; Length 456;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPOTTLSSEFENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICA 60
DB 1 GTPNPNPNPOTTLSSEFENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICA 60
QY 61 SSASYINPEGPONPNYLVTPELSLPGNGTTLFWCAQADANYASEHYAVYASSTGNDASN 120
DB 61 SSASYINPEGPONPNYLVTPELSLPGNGTTLFWCAQADANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEEVLTAKTVVTAPAIRGRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 180
DB 121 FANALLEEVLTAKTVVTAPAIRGRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 180
QY 181 DVEIKANGKRAADFTETFESETHGEAPAEWTTIDADGQGWLCSSQLGWLTAHGGTN 240
DB 181 DVEIKANGKRAADFTETFESETHGEAPAEWTTIDADGQGWLCSSQLGWLTAHGGTN 240
QY 181 DVEIKANGKRAADFTETFESETHGEAPAEWTTIDADGQGWLCSSQLGWLTAHGGTN 240
QY 241 VVASFWSNGMALPNPNYLISKDVTKATKYVYAVNDGPGPDHYAVMISKTGTDNAGDFTV 300
DB 241 VVASFWSNGMALPNPNYLISKDVTKATKYVYAVNDGPGPDHYAVMISKTGTDNAGDFTV 300
QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLNYILL 360
DB 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLNYILL 360
QY 361 DDIOFTMGSSPTPTDYYTVYVRDGTGKIKEGLTETTFEEDGVATGNHCEYCEVKYTAGVSP 420
DB 361 DDIOFTMGSSPTPTDYYTVYVRDGTGKIKEGLTETTFEEDGVATGNHCEYCEVKYTAGVSP 420
QY 421 KECVNVTVDPVQFNPVQNLTSAGVQKVTLKWDAPN 456
DB 421 KECVNVTVDPVQFNPVQNLTSAGVQKVTLKWDAPN 456

RESULT 2
US-08-570-311-20
; Sequence 20, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-20

Query Match
Best Local Similarity 100.0%; Score 2480; DB 2; Length 456;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPOTTLSSEFENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICA 60
DB 1 GTPNPNPNPOTTLSSEFENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICA 60
QY 61 SSASYINPEGPONPNYLVTPELSLPGNGTTLFWCAQADANYASEHYAVYASSTGNDASN 120
DB 61 SSASYINPEGPONPNYLVTPELSLPGNGTTLFWCAQADANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEEVLTAKTVVTAPAIRGRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 180
DB 121 FANALLEEVLTAKTVVTAPAIRGRVQGTWYQKTQVLPAGTKYVAFRHFCTDFFWNL 180
QY 181 DVEIKANGKRAADFTETFESETHGEAPAEWTTIDADGQGWLCSSQLGWLTAHGGTN 240
DB 181 DVEIKANGKRAADFTETFESETHGEAPAEWTTIDADGQGWLCSSQLGWLTAHGGTN 240
QY 241 VVASFWSNGMALPNPNYLISKDVTKATKYVYAVNDGPGPDHYAVMISKTGTDNAGDFTV 300
DB 241 VVASFWSNGMALPNPNYLISKDVTKATKYVYAVNDGPGPDHYAVMISKTGTDNAGDFTV 300
QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLNYILL 360
DB 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLNYILL 360
QY 361 DDIOFTMGSSPTPTDYYTVYVRDGTGKIKEGLTETTFEEDGVATGNHCEYCEVKYTAGVSP 420
DB 361 DDIOFTMGSSPTPTDYYTVYVRDGTGKIKEGLTETTFEEDGVATGNHCEYCEVKYTAGVSP 420
QY 421 KECVNVTVDPVQFNPVQNLTSAGVQKVTLKWDAPN 456
DB 421 KECVNVTVDPVQFNPVQNLTSAGVQKVTLKWDAPN 456

RESULT 3
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn

```

APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-14

Query Match 100.0%; Score 2480; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 3.8e-203;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTTILSEFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 60
Db 950 GTPNPNPNPGTTTILSEFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICA 1009

QY 61 SSASYINFEQPQPNPNLYLTPPELSPNGGTLTWVCAQDANYASEHYAVYASSTGNDASN 120
Db 1010 SSASYINFEQPQPNPNLYLTPPELSPNGGTLTWVCAQDANYASEHYAVYASSTGNDASN 1069

QY 121 FANALLEEVLTAKTVTVAPEAIRGTRVQGWYQKTVQLPAGTKYVAFRHFPGCTDFWNL 180
Db 1070 FANALLEEVLTAKTVTVAPEAIRGTRVQGWYQKTVQLPAGTKYVAFRHFPGCTDFWNL 1129

QY 181 DVEIKANGKRAFTETFEESSTGEPAEWTTIDADGGGWLCLSSGQLGMLTAHGNTN 240
Db 1130 DVEIKANGKRAFTETFEESSTGEPAEWTTIDADGGGWLCLSSGQLGMLTAHGNTN 1189

QY 241 VVASFSNGMALPNPNLYLISKDVTGATKVIYAVNDGFFGDHYAVMISKTGNAGDFTV 300
Db 1190 VVASFSNGMALPNPNLYLISKDVTGATKVIYAVNDGFFGDHYAVMISKTGNAGDFTV 1249

QY 301 VFETENGKNGARFGLSTEADGAKPQSWIERTVDLPAGTKYVAFRHNCSDLNYILL 360
Db 1250 VFETENGKNGARFGLSTEADGAKPQSWIERTVDLPAGTKYVAFRHNCSDLNYILL 1309

QY 361 DDIOFTMGSPPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSP 420
Db 1310 DDIOFTMGSPPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSP 1369

QY 421 KECNVNVTDPVQFNPVQNLGTSAVGQKVTLKWDAPN 456
Db 1370 KECNVNVTDPVQFNPVQNLGTSAVGQKVTLKWDAPN 1405

RESULT 4

US-08-570-311-16
Sequence 16, Application US/08570311
Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-16

Query Match 98.2%; Score 2436; DB 2; Length 450;
Best Local Similarity 99.8%; Pred. No. 1.6e-200;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 PNPNGTTLTSSFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASSASYI 66
Db 1 PNPNGTTLTSSFENGIPASWKTIDADGNNWTTTPPGGTSFAGHNSAICASSASYI 60

QY 67 NFEFGPNPDNLYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASFNALL 126
DB 61 NFEFGPNPDNLYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASFNALL 120
QY 127 BEVLTAKTVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFSGTDFFWINLDDVEIK 186
DB 121 BEVLTAKTVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFSGTDFFWINLDDVEIK 180
QY 187 ANGKRAADFTETFEFSTHGEAPAEWTTIDADGGGWLCLSSGQDLWLTAGHGTNVASFS 246
DB 181 ANGKRAADFTETFEFSTHGEAPAEWTTIDADGGGWLCLSSGQDLWLTAGHGTNVASFS 240
QY 247 WNGMALNPDNLYLSKDVGTATKYKYVAVNDGFGPDHYAVMISKTGTNAGDFTVVFETP 306
DB 241 WNGMALNPDNLYLSKDVGTATKYKYVAVNDGFGPDHYAVMISKTGTNAGDFTVVFETP 300
QY 307 NGINKGARFGLSTEADGAKPQSVWIERTVLDPAGTKYVAFRHYNCSDNLVILLDDIOFT 366
DB 301 NGINKGARFGLSTEADGAKPQSVWIERTVLDPAGTKYVAFRHYNCSDNLVILLDDIOFT 360
QY 367 MGSPTPTDYTYTVYRDGTKEGLTETTFEEDGVATGNHHEYCVVEVKYTAGVSPKBCNV 426
DB 361 MGSPTPTDYTYTVYRDGTKEGLTETTFEEDGVATGNHHEYCVVEVKYTAGVSPKBCNV 420
QY 427 TVDPVQFNPVQNLTGSAVGQKVTWKWDAPN 456
DB 421 TVDPVQFNPVQNLTGSAVGQKVTWKWDAPN 450

RESULT 5

US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguis-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 439 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-22

Query Match 94.8%; Score 2351; DB 2; Length 439;
Best Local Similarity 98.2%; Pred. No. 3e-193;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTPNPNPNPFGTTLTSESFENGIPASWKTIIDADGDNNTTTTPPGGTGTFAGHNSAICA 60
DB 1 GTPNPNPNPFGTTLTSESFENGIPASWKTIIDADGDNNTTTTPPGGTGTFAGHNSAICA 60
QY 61 SSASYINFEGPNPDNLYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
DB 61 SSASYINFEGPNPDNLYLTPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEVLTAKTVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFSGTDFFWINL 180
DB 121 FANALLEVLTAKTVTAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFSGTDFFWINL 180
QY 181 DDVEIKANGKRAADFTETFEFSTHGEAPAEWTTIDADGGGWLCLSSGQDLWLTAGHGTN 240
DB 181 DDVEIKANGKRAADFTETFEFSTHGEAPAEWTTIDADGGGWLCLSSGQDLWLTAGHGTN 240
QY 241 VVASFWSNGMALNPDNLYLSKDVGTATKYKYVAVNDGFGPDHYAVMISKTGTNAGDFTV 300
DB 241 VVASFWSNGMALNPDNLYLSKDVGTATKYKYVAVNDGFGPDHYAVMISKTGTNAGDFTV 300
QY 301 VFETTPNGINKGARFGLSTEADGAKPQSVWIERTVLDPAGTKYVAFRHYNCSDNLVILL 360
DB 301 VFETTPNGINKGARFGLSTEADGAKPQSVWIERTVLDPAGTKYVAFRHYNCSDNLVILL 360
QY 361 DDIOFTMGGSPTPTDYTYTVYRDGTKEGLTETTFEEDGVATGNHHEYCVVEVKYTAGVSP 420
DB 361 DDIOFTMGGSPTPTDYTYTVYRDGTKEGLTETTFEEDGVATGNHHEYCVVEVKYTAGVSP 420
QY 421 KECVNVTVDPVQFNPVQNL 439
DB 421 KCVNVTVINPTQFNPVQNL 439

RESULT 6

US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-09-482-500A-1

Query Match 84.0%; Score 2082; DB 4; Length 1477;

```

Best Local Similarity 86.2%; Pred. No. 2.1e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTSLSESFENGIPASWKTIDADGNNWTTTPPGCTSFAGHNSAICASS 62
Db 725 PNPENPNPGTTTSLSESFENGIPASWKTIDADGNNWTTTPPGCTSFAGHNSAICASS 781
QY 63 ASY-INFEQPNPDNYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 782 ESFGLGGIGVLTDPNLTLPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841
QY 122 ANALLEEVLTAKTVTVAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFCTDFWNLID 181
Db 842 TNALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYDLD 899
QY 182 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWLCSSGQGLWLTAGHGTNV 241
Db 900 EVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWLCSSGQGLWLTAGHGTNV 959
QY 242 VASFSWNGMALNPNDNYLISKDVTGATKYYAVYVNDGFGPDGHYAVMI-SKTGTNAGDFTVV 301
Db 960 VASFSWNGMALNPNDNYLISKDVTGATKYYAVYVNDGFGPDGHYAVMI-SKTGTNAGDFTVV 1019
QY 302 FEETPNKGGARFGLSTEADGAKPQSVWIERVTLDPAGTKYVAFRHYNCSDLYILLD 361
Db 1020 FEETPNKGGARFGLSTEADGAKPQSVWIERVTLDPAGTKYVAFRHYNCSDLYILLD 1079
QY 362 DIQFTMGGSPTDPTDYTYVYRDGKIKEGLTETTFEEDGVATGNHCVKVTAGVSPK 421
Db 1080 DIQFTMGGSPTDPTDYTYVYRDGKIKEGLTETTFEEDGVATGNHCVKVTAGVSPK 1139
QY 422 ECNVNVTDPVQFNPVQNLTSAGVQKVTWKADPN 456
Db 1140 ECNVNVTDPVQFNPVQNLTSAGVQKVTWKADPN 1174

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UFLS.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-29

Query Match 84.0%; Score 2082; DB 2; Length 1687;
Best Local Similarity 86.2%; Pred. No. 2.5e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTSLSESFENGIPASWKTIDADGNNWTTTPPGCTSFAGHNSAICASS 62
Db 935 PNPENPNPGTTTSLSESFENGIPASWKTIDADGNNWTTTPPGCTSFAGHNSAICASS 991
QY 63 ASY-INFEQPNPDNYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 992 ESFGLGGIGVLTDPNLTLPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051
QY 122 ANALLEEVLTAKTVTVAPEAIRGTRVQGTWYQKTVOLPAGTKYVAFRHFCTDFWNLID 181
Db 1052 TNALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYDLD 1109
QY 182 DVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWLCSSGQGLWLTAGHGTNV 241
Db 1110 EVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGWLCSSGQGLWLTAGHGTNV 1169
QY 242 VASFSWNGMALNPNDNYLISKDVTGATKYYAVYVNDGFGPDGHYAVMI-SKTGTNAGDFTVV 301
Db 1170 VASFSWNGMALNPNDNYLISKDVTGATKYYAVYVNDGFGPDGHYAVMI-SKTGTNAGDFTVV 1229
QY 302 FEETPNKGGARFGLSTEADGAKPQSVWIERVTLDPAGTKYVAFRHYNCSDLYILLD 361
Db 1230 FEETPNKGGARFGLSTEADGAKPQSVWIERVTLDPAGTKYVAFRHYNCSDLYILLD 1289
QY 362 DIQFTMGGSPTDPTDYTYVYRDGKIKEGLTETTFEEDGVATGNHCVKVTAGVSPK 421
Db 1290 DIQFTMGGSPTDPTDYTYVYRDGKIKEGLTETTFEEDGVATGNHCVKVTAGVSPK 1349
QY 422 ECNVNVTDPVQFNPVQNLTSAGVQKVTWKADPN 456
Db 1350 ECNVNVTDPVQFNPVQNLTSAGVQKVTWKADPN 1384

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado

```

```

; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-308A-10

Query Match 84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.6e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTIADGCGNNWTTTPPGGTSPAGHNSAICASS 62
DB 952 PNPENPNPGTTLTSESFENGIPASWKTIADGCGHGWKPGNAPG--IAGYNSNGCVYS 1008

QY 63 ASY-INFGQPNDPNVLTPELSLNGGTLTPWVCAQDANYASEHYAVYASSTGNDASNF 121
DB 1009 ESFGLGGIGVLTDPDNYLTLPALDLPNGGKLTTPWVCAQDANYASEHYAVYASSTGNDASNF 1068

QY 122 ANALLEEVLTAKTIVTAPPAIRGTRVQGTWYOKTVOLPAGTKYVAFRHPGCTDFFWINLD 181
DB 1069 TWALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHPQSTDMFYIDL 1126

QY 182 DVEIKANGKRADFTTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGGTNNV 241
DB 1127 EVEIKANGKRADFTTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDMLTAHGGTNNV 1186

QY 242 VASFSWNGMALNPDNVLISKDVTKATKYYAVNDGFGPDHAYVMIKGTGNAGDFTVV 301
DB 1187 VASFSWNGMALNPDNVLISKDVTKATKYYAVNDGFGPDHAYVMIKGTGNAGDFTVV 1246

QY 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 361
DB 1247 FEETPNKNGGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 1306

QY 362 DIQFTMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEDGVATGNHEYCEVVKYTAGVSPK 421
DB 1307 DIQFTMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEDGVATGNHEYCEVVKYTAGVSPK 1366

QY 422 ECNVNVTVDVQFNPQNLTGSAVGQKVTLLKWDAPN 456
DB 1367 ECNVNVTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1401

```

RESULT 9

US-08-822-324-6

; Sequence 6, Application US/08822324

; Patent No. 6129917

```

; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 488-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-822-324-6

```

```

Query Match 84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.6e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTIADGCGNNWTTTPPGGTSPAGHNSAICASS 62
DB 952 PNPENPNPGTTLTSESFENGIPASWKTIADGCGHGWKPGNAPG--IAGYNSNGCVYS 1008

QY 63 ASY-INFGQPNDPNVLTPELSLNGGTLTPWVCAQDANYASEHYAVYASSTGNDASNF 121
DB 1009 ESFGLGGIGVLTDPDNYLTLPALDLPNGGKLTTPWVCAQDANYASEHYAVYASSTGNDASNF 1068

QY 122 ANALLEEVLTAKTIVTAPPAIRGTRVQGTWYOKTVOLPAGTKYVAFRHPGCTDFFWINLD 181
DB 1069 TWALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHPQSTDMFYIDL 1126

QY 182 DVEIKANGKRADFTTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGGTNNV 241
DB 1127 EVEIKANGKRADFTTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDMLTAHGGTNNV 1186

QY 242 VASFSWNGMALNPDNVLISKDVTKATKYYAVNDGFGPDHAYVMIKGTGNAGDFTVV 301
DB 1187 VASFSWNGMALNPDNVLISKDVTKATKYYAVNDGFGPDHAYVMIKGTGNAGDFTVV 1246

QY 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 361
DB 1247 FEETPNKNGGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 1306

QY 362 DIQFTMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEDGVATGNHEYCEVVKYTAGVSPK 421
DB 1307 DIQFTMGGSPTPTDYTYTVYRDGTIKIKEGLTETTFEDGVATGNHEYCEVVKYTAGVSPK 1366

```



```
QY 422 ECVNVTVDVQFNPVQNLTSAGVQKVTLLKWDAPN 456
|||||:|||||:|||||:|||||:|||||:
Db 1367 ECVNVTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1401
|||||:|||||:|||||:|||||:|||||:

RESULT 10
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-490-931-10

Query Match 84.0%; Score 2082; DB 3; Length 1704;
Best Local Similarity 86.2%; Pred. No. 2.6e-169;
Matches 392; Conservative 17; Mismatches 40; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTILSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICASS 62
Db 952 PNPENPNPGTTTILSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICASS 1008
|||||:|||||:|||||:|||||:|||||:

QY 63 ASY-INFEQPNPNVLTPELSIPNGGTLTFWVCAQADANYASEHYAVYASSTGNDASNF 121
Db 1009 ESFGLGGIGVLTDPNYLITPALDLPNGGKLTFWVCAQADANYASEHYAVYASSTGNDASNF 1068
|||||:|||||:|||||:|||||:|||||:

QY 122 ANALLEEVLTAKTVVTAPEAIRGRVQGTWQKTVDLPAGTKYVAFRHFQCTDFFWINLD 181
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQCTDFFWINLD 1126
|||||:|||||:|||||:|||||:|||||:

QY 182 DVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQWLCCLSSGQLDHLTAHGNTV 241
Db 182 DVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQWLCCLSSGQLDHLTAHGNTV 241
|||||:|||||:|||||:|||||:|||||:

QY 242 VASFSWNGMALNPDNYLISKDVTGATKVKYKYAVVNDGFFPDGHYAVVMI-SKTGTNAGDFTVV 301
Db 242 VASFSWNGMALNPDNYLISKDVTGATKVKYKYAVVNDGFFPDGHYAVVMI-SKTGTNAGDFTVV 301
|||||:|||||:|||||:|||||:|||||:

QY 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 361
Db 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 361
|||||:|||||:|||||:|||||:|||||:

QY 362 DIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEVCVEVKYTAGVSPK 421
Db 362 DIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEVCVEVKYTAGVSPK 421
|||||:|||||:|||||:|||||:|||||:

Db 1127 EVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQWLCCLSSGQLDHLTAHGNTV 1186
QY 242 VASFSWNGMALNPDNYLISKDVTGATKVKYKYAVVNDGFFPDGHYAVVMI-SKTGTNAGDFTVV 301
Db 1187 VASFSWNGMALNPDNYLISKDVTGATKVKYKYAVVNDGFFPDGHYAVVMI-SKTGTNAGDFTVV 1246
QY 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 361
Db 1247 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 1306
QY 362 DIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEVCVEVKYTAGVSPK 421
Db 1307 DIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEVCVEVKYTAGVSPK 1366
QY 422 ECVNVTVDVQFNPVQNLTSAGVQKVTLLKWDAPN 456
Db 1367 ECVNVTINPTQFNPVKNLKAQPDGDDVVLKWEAPS 1401

RESULT 11
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-10

Query Match 83.5%; Score 2070; DB 4; Length 1706;
Best Local Similarity 85.7%; Pred. No. 2.7e-168;
Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTILSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICASS 62
Db 954 PNPENPNPGTTTILSEFENGIPASWKTIDADGDNWTTTPPPGGTSFAGHNSAICASS 1010
|||||:|||||:|||||:|||||:|||||:

QY 63 ASY-INFEQPNPNVLTPELSIPNGGTLTFWVCAQADANYASEHYAVYASSTGNDASNF 121
Db 1011 ESFGLGGIGVLTDPNYLITPALDLPNGGKLTFWVCAQADANYASEHYAVYASSTGNDASNF 1070
|||||:|||||:|||||:|||||:|||||:

QY 122 ANALLEEVLTAKTVVTAPEAIRGRVQGTWQKTVDLPAGTKYVAFRHFQCTDFFWINLD 181
Db 1071 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQCTDFFWINLD 1128
|||||:|||||:|||||:|||||:|||||:

QY 182 DVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQWLCCLSSGQLDHLTAHGNTV 241
Db 1129 EVEIKANGKRAADTFETPESSTHGEAPAEWTTIDADGGQWLCCLSSGQLDHLTAHGNTV 1188
|||||:|||||:|||||:|||||:|||||:

QY 242 VASFSWNGMALNPDNYLISKDVTGATKVKYKYAVVNDGFFPDGHYAVVMI-SKTGTNAGDFTVV 301
Db 1189 VSSFSWNGMALNPDNYLISKDVTGATKVKYKYAVVNDGFFPDGHYAVVMI-SKTGTNAGDFTVV 1248
|||||:|||||:|||||:|||||:|||||:

QY 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 361
Db 1249 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDNLNILLD 1308
|||||:|||||:|||||:|||||:|||||:

QY 362 DIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEVCVEVKYTAGVSPK 421
Db 362 DIQFTMGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHHEVCVEVKYTAGVSPK 421
|||||:|||||:|||||:|||||:|||||:
```

Db 1309 DIOFTMGSPPTDYTYVYRDGTIKKEGLTETTFEEDGVATGNHEVCVEKVTAGVSPK 1368

QY 422 ECNVNTPVPQFNPNVQNLTSAGVGKVTWKDAPN 456

Db 1369 KCVNVTNSTQFNPNVKNLKAQPDGGDVVLKWEAPS 1403

RESULT 12

US-08-570-311-8

; Sequence 8, Application US/08570311

; Patent No. 5824791

; GENERAL INFORMATION:

; APPLICANT: Proguiske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guylaine

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ted W. Whitlock

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/570.311

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/353,485

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/647,119

; FILING DATE: 08-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UF15.C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (904) 375-8100

; TELEFAX: (904) 372-5800

; INFORMATION FOR SEQ ID NO: 8:

; LENGTH: 1087 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-570-311-8

Query Match 83.3%; Score 2066; DB 2; Length 1087;

Best Local Similarity 85.3%; Pred No. 3, 1e-168;

Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

QY 1 GTPNPNPNPQTTLSFENGIPASWKTIDADGGNNWTTTPPPGGTSPAGHNSAICA 60

Db 333 GTPNPNPNPQTTLSFENGIPASWKTIDADGGHGWKPGNAPG--JAGYNSNGCV 389

QY 61 SSASY-IFEGQPNPDNVLVTELSLPNGGTTFFWVCAQDANYASEHYAVYASSTGNDAS 119

Db 390 YSBSFGLGGIGVLTPDNYLITPALDLANGGKLTFFWVCAQDANYASEHYAVYASSTGNDAS 449

QY 120 NFANALLEEVLTAKTVVTAPBAIRGRVQGTWYQKTQVQLPAGTKYVAFRHPGCTDFFWIN 179

Db 450 NFTNALLEETITAGK-VRSPEAIRG-RIOGTWRQKTVDLPAGTKYVAFRHPFQSTDMFYID 507

QY 180 LDDVEIKANGKRADFTETFPESSTHGEAPAEWTTTIDADGGGWLCLSSGOLGWLTAHGCT 239

Db 508 LDEVEIKANGKRADFTETFPESSTHGEAPAEWTTTIDADGGQDWLCLSSGQLDWTAGGT 567

QY 240 NVASFSWNGWALNPDNYLISKDVTGATKVKYVYVANDPFGDHYAVMIKSTGNAGDFT 299

Db 568 NVASFSWNGWALNPDNYLISKDVTGATKVKYVYVANDPFGDHYAVMIKSTGNAGDFT 627

QY 300 VFEETENGINKGARFGLSTEADGAKPQSVWERTVLDLPAGTKYVAFRHYNCSDLVIL 359

Db 628 VFEETENGINKGARFGLSTEADGAKPQSVWERTVLDLPAGTKYVAFRHYNCSDLVIL 687

QY 360 LDDIQFTMGSPPTDYTYVYRDGTIKKEGLTETTFEEDGVATGNHEVCVEKVTAGVS 419

Db 688 LDDIQFTMGSPPTDYTYVYRDGTIKKEGLTETTFEEDGVATGNHEVCVEKVTAGVS 747

QY 420 PKCVNTPVPQFNPNVQNLTSAGVGKVTWKDAPN 456

Db 748 PKCVNTPVPQFNPNVKNLKAQPDGGDVVLKWEAPS 784

RESULT 13

US-08-353-485-8

; Sequence 8, Application US/08353485

; Patent No. 5830710

; GENERAL INFORMATION:

; APPLICANT: Proguiske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guylaine

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ted W. Whitlock

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/353,485

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/647,119

; FILING DATE: 25-JAN-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/241,640

; FILING DATE: 08-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Whitlock, Ted W.

; REGISTRATION NUMBER: 36,965

; REFERENCE/DOCKET NUMBER: UF15.C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (904) 375-8100

; TELEFAX: (904) 372-5800

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

```
/
/ LENGTH: 1087 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-08-353-485-8
/
Query Match
Best Local Similarity 83.3%; Score 2066; DB 2; Length 1087;
Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

QY 1 GTPNPNPNPGTTLSSEFENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICA 60
Db 333 GTPNPNPNPGTTLSSEFENGIPASWKTIADGDGHCWKPGNAPG---IAGYNSGCV 389
/
QY 61 SSASY-INPEGQNPNDYLVTPELSLPNGGTLTFWVCAODANVASEHYAVYASSTGNDAS 119
Db 390 YSEFGLGIGVLTDPNYLITPALDLANGKLTFFWCAODANVASEHYAVYASSTGNDAS 449
/
QY 120 NFANALLEEVLTAKTVVTAPAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWIN 179
Db 450 NFNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAAGTKYVAFRHFQSTDMFYID 507
/
QY 180 LDDVEIKANGKRAADFTTETFESETHGEPAPAEWTTIDADGCGQWLCISSLGQLDMLTAHGGT 239
Db 508 LDEVEIKANGKRAADFTTETFESETHGEPAPAEWTTIDADGCGQWLCISSLGQLDMLTAHGGT 567
/
QY 240 NVVASFWSWGMALNPNDYLSKDVTCATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFT 299
Db 568 NVVASFWSWGMALNPNDYLSKDVTCATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFT 627
/
QY 300 VVFEETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 359
Db 628 VVFEETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 687
/
QY 360 LDDIQTWCGSPPTDYYTVYRDGTHKKEGLTETTFEEDGVATGNHCVKVTAGVS 419
Db 688 LDDIQTWCGSPPTDYYTVYRDGTHKKEGLTETTFEEDGVATGNHCVKVTAGVS 747
/
QY 420 PKECVNVTVDPVQFNPVQNLTGSAVGOKVTLKWDAPN 456
Db 748 PKVCVNVNTINPTQFNPVKNLKAQPDGDDVVVKWEAPS 784
/
RESULT 14
US-08-570-311-27
/ Sequence 27, Application US/08570311
/ Patent No. 5824791
/
GENERAL INFORMATION:
/ APPLICANT: Proguiske-Fox, Ann
/ APPLICANT: Tunwasorn, Somying
/ APPLICANT: Lepine, Guylaine
/ APPLICANT: Han, Naiming
/ APPLICANT: Lantz, Marilyn
/ APPLICANT: Patti, Joseph
/ TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
/ TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
/ NUMBER OF SEQUENCES: 29
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Ted W. Whitlock
/ STREET: 2421 N.W. 41st Street, Suite A-1
/ CITY: Gainesville
/ STATE: FL
/ COUNTRY: USA
/ ZIP: 32606
/
COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/570,311
/ FILING DATE:
/ CLASSIFICATION: 424
/
PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 08/353,485
/ FILING DATE: 09-DEC-1994
/ CLASSIFICATION: 424
/
PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 07/647,119
/ FILING DATE: 25-JAN-1991
/ CLASSIFICATION: 424
/
PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 07/241,640
/ FILING DATE: 08-SEP-1988
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Whitlock, Ted W.
/ REGISTRATION NUMBER: 36,965
/ REFERENCE/DOCKET NUMBER: UFI5.C3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (904) 375-8100
/ TELEFAX: (904) 372-5800
/
INFORMATION FOR SEQ ID NO: 27:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1358 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-08-570-311-27
/
Query Match
Best Local Similarity 83.3%; Score 2066; DB 2; Length 1358;
Matches 390; Conservative 18; Mismatches 43; Indels 6; Gaps 4;

QY 1 GTPNPNPNPGTTLSSEFENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICA 60
Db 604 GTPNPNPNPGTTLSSEFENGIPASWKTIADGDGHCWKPGNAPG---IAGYNSGCV 660
/
QY 61 SSASY-INPEGQNPNDYLVTPELSLPNGGTLTFWVCAODANVASEHYAVYASSTGNDAS 119
Db 661 YSEFGLGIGVLTDPNYLITPALDLANGKLTFFWCAODANVASEHYAVYASSTGNDAS 720
/
QY 120 NFANALLEEVLTAKTVVTAPAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWIN 179
Db 721 NFNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAAGTKYVAFRHFQSTDMFYID 778
/
QY 180 LDDVEIKANGKRAADFTTETFESETHGEPAPAEWTTIDADGCGQWLCISSLGQLDMLTAHGGT 239
Db 779 LDEVEIKANGKRAADFTTETFESETHGEPAPAEWTTIDADGCGQWLCISSLGQLDMLTAHGGT 838
/
QY 240 NVVASFWSWGMALNPNDYLSKDVTCATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFT 299
Db 839 NVVASFWSWGMALNPNDYLSKDVTCATKVKYVAVNDGFPDGHYAVMISKTGTNAGDFT 898
/
QY 300 VVFEETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 359
Db 899 VVFEETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIL 958
/
QY 360 LDDIQTWCGSPPTDYYTVYRDGTHKKEGLTETTFEEDGVATGNHCVKVTAGVS 419
Db 959 LDDIQTWCGSPPTDYYTVYRDGTHKKEGLTETTFEEDGVATGNHCVKVTAGVS 1018
/
QY 420 PKECVNVTVDPVQFNPVQNLTGSAVGOKVTLKWDAPN 456
Db 1019 PKVCVNVNTINPTQFNPVKNLKAQPDGDDVVVKWEAPS 1055
/
RESULT 15
US-09-066-330-11
/ Sequence 11, Application US/09066330A
/ Patent No. 6511666
/
GENERAL INFORMATION:
/ APPLICANT: Reynolds, Eric C.
/ APPLICANT: Bhogal, Peter S.
/ APPLICANT: Slakeski, Nada
/ TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
/ FILE REFERENCE: Reynolds
```

```
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match      82.2%; Score 2039.5; DB 4; Length 1732;
Best Local Similarity 84.5%; Pred.No. 1.2e-165;
Matches 388; Conservative 18; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPQT-TTILSESPENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAIC 59
Db 969 GTPNPNPNPNPQTTLSESPENGIPASWKTIADGDGHWKPGNAPG--IAGYNSGC 1025
QY 60 ASSASY-INFGPQNDNLYVTPELSLNGGTLTFWVCAQDANYASEHYAVYASGTGND 118
Db 1026 VYSEFGLGGIGVLTDPDNLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASGTGND 1085
QY 119 SNEFALLLEVLITAKTVTAPETVQGTWYQKTVOLPAGTKYVAFRHEGCTDFFWI 178
Db 1086 SNFTNALLEETITAKG-VRSKPAIRG-RIQGTWROKTVLDPAGTKYVAFRHFQSDMEFY 1143
QY 179 NLDDVEIKANGKRADETFTESSHGEAPAEWTTIADGDGQGLCLSSGQLGWLTAHGG 238
Db 1144 DLDEVEIKANGKRADETFTESSHGEAPAEWTTIADGDGQGLCLSSGQLDMLTAHGG 1203
QY 239 TNVASFNGMGNALNPDNYLISKDVTKYKYAVNDGPPGDHYAVMISKTGTNAGDF 298
Db 1204 SNVSSFNGMGNALNPDNYLISKDVTKYKYAVNDGPPGDHYAVMISKTGTNAGDF 1263
QY 299 TVVFEETPNKNGGARFGLSTEADGAKPQSVWIERTVLDPAGTKYVAFRHYNCSDLNYI 358
Db 1264 TVVFEETPNKNGGARFGLSTEADGAKPQSVWIERTVLDPAGTKYVAFRHYNCSDLNYI 1323
QY 359 LDDDIQFTMGGSPTPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGV 418
Db 1324 LDDDIQFTMGGSPTPTDYTYTVYRDGKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGV 1383
QY 419 SPKCVNVTVPQNPVQNLGSAV--GQKVTLKWDAP 455
Db 1384 SPKCVNVTVNSTQNPVQNLTAEOAPNSMDAILKWNAP 1422
```

Search completed: May 18, 2004, 11:49:16
Job time : 16.0888 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:42:50 ; Search time 35.3211 Seconds
(without alignments)

3592.387 Million cell updates/sec

Title: US-08-570-311-20

Perfect score: 2480

Sequence: 1 GTPNPNPNGPCTTILSESF.....QNLTSVAGQKVTILKWDAPN 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--|
| 1 | 2070 | 83.5 | 1706 | 14 | US-10-229-066-10 Sequence 10, Appl |
| 2 | 2039.5 | 82.2 | 1732 | 14 | US-10-229-066-11 Sequence 11, Appl |
| 3 | 726 | 29.3 | 419 | 15 | US-10-174-695-3 Sequence 3, Appl |
| 4 | 710.5 | 28.6 | 419 | 15 | US-10-174-695-5 Sequence 5, Appl |
| 5 | 385 | 15.5 | 231 | 15 | US-10-174-695-6 Sequence 6, Appl |
| 6 | 232.5 | 9.4 | 196 | 15 | US-10-174-695-4 Sequence 4, Appl |
| 7 | 153 | 6.2 | 29 | 15 | US-10-387-977-18 Sequence 18, Appl |
| 8 | 153 | 6.2 | 29 | 15 | US-10-387-977-21 Sequence 21, Appl |
| 9 | 153 | 6.2 | 29 | 15 | US-10-387-977-23 Sequence 23, Appl |
| 10 | 153 | 6.2 | 29 | 15 | US-10-387-977-24 Sequence 24, Appl |
| 11 | 153 | 6.2 | 29 | 15 | US-10-387-977-26 Sequence 26, Appl |
| 12 | 147 | 5.9 | 29 | 15 | US-10-387-977-17 Sequence 17, Appl |
| 13 | 147 | 5.9 | 872 | 12 | US-10-282-122A-55467 Sequence 55467, A |
| 14 | 145 | 5.8 | 29 | 15 | US-10-387-977-25 Sequence 25, Appl |
| 15 | 143 | 5.8 | 509 | 15 | US-10-387-977-101 Sequence 101, Appl |

| | | | | | | |
|----|-------|-----|------|----|----------------------|-------------------|
| 16 | 137 | 5.5 | 1833 | 14 | US-10-175-282-4 | Sequence 4, Appl |
| 17 | 137 | 5.5 | 1833 | 14 | US-10-175-282-4 | Sequence 4, Appl |
| 18 | 137 | 5.5 | 1992 | 14 | US-10-175-282-3 | Sequence 3, Appl |
| 19 | 137 | 5.5 | 1992 | 14 | US-10-175-282-3 | Sequence 3, Appl |
| 20 | 136 | 5.5 | 29 | 15 | US-10-387-977-20 | Sequence 20, Appl |
| 21 | 134 | 5.4 | 25 | 15 | US-10-387-977-77 | Sequence 77, Appl |
| 22 | 134 | 5.4 | 25 | 15 | US-10-387-977-80 | Sequence 80, Appl |
| 23 | 134 | 5.4 | 27 | 15 | US-10-387-977-11 | Sequence 11, Appl |
| 24 | 134 | 5.4 | 27 | 15 | US-10-387-977-12 | Sequence 12, Appl |
| 25 | 134 | 5.4 | 27 | 15 | US-10-387-977-13 | Sequence 13, Appl |
| 26 | 132.5 | 5.3 | 555 | 14 | US-10-156-761-14592 | Sequence 14592, A |
| 27 | 132.5 | 5.3 | 691 | 15 | US-10-369-493-23473 | Sequence 23473, A |
| 28 | 131.5 | 5.3 | 2122 | 9 | US-09-813-214A-9 | Sequence 9, Appl |
| 29 | 131.5 | 5.3 | 2468 | 12 | US-10-282-122A-66335 | Sequence 66335, A |
| 30 | 131.5 | 5.3 | 2468 | 14 | US-10-246-330-4 | Sequence 4, Appl |
| 31 | 131 | 5.3 | 960 | 14 | US-10-311-879-15 | Sequence 15, Appl |
| 32 | 130 | 5.2 | 25 | 15 | US-10-387-977-81 | Sequence 81, Appl |
| 33 | 130 | 5.2 | 1946 | 12 | US-10-282-122A-62947 | Sequence 62947, A |
| 34 | 129.5 | 5.2 | 1638 | 12 | US-10-206-576-258 | Sequence 258, App |
| 35 | 129.5 | 5.2 | 1638 | 12 | US-10-206-576-262 | Sequence 262, App |
| 36 | 129.5 | 5.2 | 1638 | 12 | US-10-206-576-266 | Sequence 266, App |
| 37 | 129.5 | 5.2 | 1728 | 12 | US-10-282-122A-56997 | Sequence 56997, A |
| 38 | 129 | 5.2 | 24 | 14 | US-10-229-066-6 | Sequence 6, Appl |
| 39 | 129 | 5.2 | 2435 | 12 | US-10-282-122A-47453 | Sequence 47453, A |
| 40 | 127 | 5.1 | 25 | 15 | US-10-387-977-83 | Sequence 83, Appl |
| 41 | 127 | 5.1 | 25 | 15 | US-10-387-977-84 | Sequence 84, Appl |
| 42 | 127 | 5.1 | 25 | 15 | US-10-387-977-85 | Sequence 85, Appl |
| 43 | 126 | 5.1 | 1649 | 15 | US-10-369-493-18460 | Sequence 18460, A |
| 44 | 123.5 | 5.0 | 491 | 15 | US-10-387-977-100 | Sequence 100, App |
| 45 | 123.5 | 5.0 | 669 | 12 | US-10-206-576-264 | Sequence 264, App |

ALIGNMENTS

RESULT 1

US-10-229-066-10
; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 83.5%; Score 2070; DB 14; Length 1706;
Best Local Similarity 85.7%; Pred. No. 3.9e-176;
Matches 390; Conservative 18; Mismatches 41; Indels 6; Gaps 4;

QY 3 PNPNPNPCTTILSESFENGIPASWKITDADGNNWTTTPPPGGTSPAGHNSAICASS 62
|||
DB 954 PNPNPNPCTTILSESFENGIPASWKITDADGNGKPKGNAPG---TAGYNSNGCVYS 1010
|||
QY 63 ASY-INFEQPNPDNYLVTPELSLNGGTLTFWVCAQDANYASEHVAVYASSTGNDASNF 121
|||
DB 1011 ESFGLGIGVLTDPDNYLITPALDPNGGKLTFWVCAQDANYASEHVAVYASSTGNDASNF 1070
|||

```
QY 122 ANALLEEVLTAKTVTVAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFWNLND 181
Db 1071 TNALLEETITAKG-VRSPEAKRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDL 1128
QY 182 DVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGGQGWHLCLSSGQLWLTAGGTNV 241
Db 1129 EVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGGQGWHLCLSSGQLWLTAGGTNV 1188
QY 242 VASFSWNGMALNPNDNLYLSKDVATKVKYKYAVNDGPPGDHYAVMISKTGTNAGDFTV 301
Db 1189 VSSFSWNGMALNPNDNLYLSKDVATKVKYKYAVNDGPPGDHYAVMISKTGTNAGDFTV 1248
QY 302 FEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLNLYILL 361
Db 1249 FEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLNLYILL 1308
QY 362 DIQFTMGSSPTDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHCVKVTAGVSPK 421
Db 1309 DIQFTMGSSPTDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHCVKVTAGVSPK 1368
QY 422 ECVNVTVDPQFNPVQNLTSAGVGOKVTLKWDAPN 456
Db 1369 KCVNVTVNSTQFNPVQNLKXQPDGDDVVLKWEAPS 1403

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match 82.2%; Score 2039.5; DB 14; Length 1732;
Best Local Similarity 84.5%; Pred. No. 2.2e-173;
Matches 388; Conservative 18; Mismatches 44; Indels 9; Gaps 6;

QY 1 GTPNPNPNPGT-TTLLSEFNGIPASWKTIADADGNNWTTTPPGGTSPAGHNSAIC 59
Db 969 GTPNPNPNPNPGT-TTLLSEFNGIPASWKTIADADGNNWTTTPPGGTSPAGHNSAIC 1025
QY 60 ASSASY-INFEQPQNPONLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGND 118
Db 1026 VYSEFGLGGIGVLTTPDNLYITPALDPLNGKLTFFWVCAQDANYASEHYAVYASSTGND 1085
QY 119 SNFANALLEEVLTAKTVTVAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFW 178
Db 1086 SNFTNALLEETITAKG-VRSPEAKRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYI 1143
QY 179 NLDDVEIKANGKRAADFTETTESSTHGEAPAEWTITDADGGQGWHLCLSSGQLWLTAG 238
Db 1144 DLDEVIKANGKRAADFTETTESSTHGEAPAEWTITDADGGQGWHLCLSSGQLWLTAG 1203
QY 239 TNVVASFSWNGMALNPNDNLYLSKDVATKVKYKYAVNDGPPGDHYAVMISKTGTNAGD 298
Db 1144 DLDEVIKANGKRAADFTETTESSTHGEAPAEWTITDADGGQGWHLCLSSGQLWLTAG 1203
```

```
Db 1204 SNVVSFSWNGMALNPNDNLYLSKDVATKVKYKYAVNDGPPGDHYAVMISKTGTNAGD 1263
QY 299 TVVFEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLN 358
Db 1264 TVVFEETPNGINKGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNCSDLN 1323
QY 359 LLDDIQFTMGSSPTDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHCVKVTAGV 418
Db 1324 LLDDIQFTMGSSPTDYTYTVYRDGKIKIEGLTETTFEEDGVATGNHCVKVTAGV 1383
QY 419 SPKECVNVTVDPQFNPVQNLTSAGV--GQKVTLKWDAP 455
Db 1384 SPKCVNVTVNSTQFNPVQNLTAEOAPNSMDAILKWNAP 1422

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match 29.3%; Score 726; DB 15; Length 419;
Best Local Similarity 75.3%; Pred. No. 2.3e-56;
Matches 143; Conservative 11; Mismatches 30; Indels 6; Gaps 4;

QY 3 PNPENPNPNPGT-TTLLSEFNGIPASWKTIADADGNNWTTTPPGGTSPAGHNSAIC 62
Db 235 PNPENPNPNPGT-TTLLSEFNGIPASWKTIADADGNNWTTTPPGGTSPAGHNSAIC 291
QY 63 ASY-INFEQPQNPONLVTPPELSPNGGTLTFWVCAQDANYASEHYAVYASSTGNDAGNF 121
Db 292 ESFGUGGIGVLTTPDNLYITPALDPLNGKLTFFWVCAQDANYASEHYAVYASSTGNDAGNF 351
QY 122 ANALLEEVLTAKTVTVAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFWNLND 181
Db 352 TNALLEETITAKG-VRSPEAKRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDL 409
QY 182 DVEIKANGKR 191
Db 410 EVEIKANGKR 419

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
```

```

; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-5
Query Match 28.6%; Score 710.5; DB 15; Length 419;
Best Local Similarity 73.6%; Pred. No. 5.6e-55;
Matches 142; Conservative 12; Mismatches 32; Indels 7; Gaps 5;

Qy 1 GTPNPNPNPNT-TTILSEFENGIPASWKTTDADGNNWTTTPPGTSPAGHNSAIC 59
Db 232 GTPNPNPNPNT-TTILSEFENGIPASWKTTDADGNGKPGNAPG---IAGYNSGC 288
Qy 60 ASSASY-INFEQPNPNYLVTPPELSLPGNGTLTFWVCAQDANYASEHYAVASSTGND 118
Db 289 VYESFGGIGIVLPEDNYLITLALDPLNGKLTTFWVCAQDANYASEHYAVASSTGND 348
Qy 119 SNFANALLEVLTAKTVTVTAPAIRTRVGQTYQKTVQLPAGTKYVAFRHFQCTDFWI 178
Db 349 SNFTNALLEETITAKG-VRSPKAIIRG-RIQGTWRQKTVLPAGTKYVAFRHFQSTDMFVI 406
Qy 179 NLDDVEIKANGKR 191
Db 407 DLDEVEIKANGKR 419

RESULT 5
US-10-174-695-6
; Sequence 6, Application US/10174695
; Publication No. US2003023022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-4
Query Match 9.4%; Score 232.5; DB 15; Length 196;
Best Local Similarity 39.1%; Pred. No. 1.5e-12;
Matches 61; Conservative 13; Mismatches 63; Indels 19; Gaps 5

Qy 3 PNPNPNPCTTTLSSEFENGIPASWKTTDADGNNWTTTPPGTSPAGHNSAICASS 62
Db 44 PNPNPNPCTTTLSSEFENGIPASWKTTDADGNGKPGNAPG---IAGYNSGCVYL 100
Qy 63 ASYNFEGPQPNYLVTPPELSLPGNGTLTFWVCAQDANYASEHYAVASSTGNDASNFA 122
Db 101 DNSAKIDRNCQEIINVY-NTAEYAKTNNAPI-----KVVGAYDE-----KTGTAAVNN- 145
Qy 123 NALLEEVLTAKTVTVTAPAIRTRVGQTYQKTVQL 158
Db 146 -KLSERRAKAVAKMLEKYGVSDRITIEWKSSEQI 180

RESULT 7
US-10-387-977-18
; Sequence 18, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PQ 6528
; PRIOR FILING DATE: 1997-04-30

```

```
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

Query Match          6.2%  Score 153;  DB 15;  Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGKIKKEGLTETTFEEDGVAT 403
    |||||
Db 1 DYTIVYRDGKIKKEGLTETTFEEDGVAT 29

RESULT 8
US-10-387-977-21
; Sequence 21, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-21

Query Match          6.2%  Score 153;  DB 15;  Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGKIKKEGLTETTFEEDGVAT 403
    |||||
Db 1 DYTIVYRDGKIKKEGLTETTFEEDGVAT 29

RESULT 9
US-10-387-977-23
; Sequence 23, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-23

Query Match          6.2%  Score 153;  DB 15;  Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGKIKKEGLTETTFEEDGVAT 403
    |||||
Db 1 DYTIVYRDGKIKKEGLTETTFEEDGVAT 29

RESULT 10
US-10-387-977-24
; Sequence 24, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-24

Query Match          6.2%  Score 153;  DB 15;  Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGKIKKEGLTETTFEEDGVAT 403
    |||||
Db 1 DYTIVYRDGKIKKEGLTETTFEEDGVAT 29

RESULT 11
US-10-387-977-26
; Sequence 26, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match          6.2%  Score 153;  DB 15;  Length 29;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGKIKKEGLTETTFEEDGVAT 403
    |||||
Db 1 DYTIVYRDGKIKKEGLTETTFEEDGVAT 29
```



```

; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-26

Query Match
Best Local Similarity 6.2%; Score 153; DB 15; Length 29;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTIKKEGLTTFEEDGVAT 403
Db 1 DYTIVYRDGTIKKEGLTTFEEDGVAT 29

RESULT 12
US-10-387-977-17
; Sequence 17, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; FILE REFERENCE: PORPHYROMONAS GINGIVALIS
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-17

Query Match
Best Local Similarity 5.9%; Score 147; DB 15; Length 29;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTIKKEGLTTFEEDGVAT 403
Db 1 DYTIVYRDGTIKKEGLTTFEEDGVAT 29

RESULT 13
US-10-282-122A-55467
; Sequence 55467, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55467
; LENGTH: 872
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-55467

Query Match
Best Local Similarity 5.9%; Score 147; DB 12; Length 872;
Matches 112; Conservative 56; Mismatches 205; Indels 154; Gaps 23;

QY 10 NFGTTTLESFENGIPASWKTIDADGDGNNWTTTTPPGGTSPAGHNSAICASSASYNFE 69
Db 293 NTGTGTHNVTVDTGLPS--VAFNAISDDNVLNAVEKGQDLSVSGTSANLAEGTVTVTLN 350
QY 70 GPQNPNDVLTPELSLPNGGTLTFWVCAQDANVASEHYAVYASST---GNDASNFANALL 126
Db 351 G-----KNYAAATTAAD---GTWSLTVPAADLAGLGHDTLSATATNGVGSNTANLLV 402
QY 127 BEVLTAKT-----VVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWI 178
Db 403 DTALPTVTINTVAGDUNVINAAEVAAGQTISG-----KVANAEG----- 441
QY 179 NLDDVEIKANGKRAPTTFETSTHGEAPAEWTTIDADGGQGWLCSSGQLGWLTAHGG 238
Db 442 NTVTVTIGN-----SYTATVQSDLTWSVNVPSVLTALNGD--LTVSATV---TNGHGN 492
QY 239 T-----NVVASPSWNGMALNPNDYLISKDVTGATKVKYVAVYVNDGPPGHHYAMISKTGT 293
Db 493 TGAGEREIVIDASLPGLRIDT---VAGDDV-----INSIEHQNLIVTSSDGL 538
QY 294 NAGDFTVFEETPNGINKGGAFLSTEADGAKPQSVWITERTVDLPAGT----- 342
Db 539 AAG-----TTLTVVNGKTYAASVLADGTWSAIPAADYCALAAGTIVTTAAGQSAA 590
QY 343 -KYVAFRHYNCSDNLNILL-----DDI-----QFTMGGSPPPTD----- 375
Db 591 GNPVTISHDVTVDLSAVAISIDAIAETDDVINAEEKGADLVLSGTSNVEENQTVTITSGG 650
QY 376 --YTYTVYEDG-----TKIKEG---LTETTFEEDGVATCNHVCYEVKVTAGV 418
Db 651 KTYTAKVDADGNWTATVPSADLAGLKGDDASVQVSVTNAHNSASASAGREYSVD-----AT 705
QY 419 SPKECVNVTVDVPQVFNPNQNLTGSA-----VGQKVTWKWD 453
Db 706 AP-----TVTIDTVAGDNNVINASEAAAGVAISGTTTAEVGGTIVTVILD 748

```

RESULT 14
 US-10-387-977-25
 ; Sequence 25, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 25
 ; LENGTH: 29
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-25

Query Match 5.8%; Score 145; DB 15; Length 29;
 Best Local Similarity 96.6%; Pred. No. 6.6e-06;
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 375 DYTIVYRDGKIKKEGLTETTFEEDGVAT 403
 Db 1 DYTIVYRDGKIKRWGLTETTFEEDGVAT 29

RESULT 15
 US-10-387-977-101
 ; Sequence 101, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 101
 ; LENGTH: 509
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-101

Query Match 5.8%; Score 143; DB 15; Length 509;
 Best Local Similarity 20.0%; Pred. No. 0.00066;
 Matches 94; Conservative 66; Mismatches 189; Indels 120; Gaps 21;

Qy 54 HNSAICASAS-YINFEQPNPNLYVTPSLPNSGTLTFWVCAQDANYASEHYAV-YA 111
 Db 67 YNDGLAASAPVFLALVG---DTDVISGEKGGKTKKVTDLYSVDGDFPEMYTFRMS 122

Qy 112 SSTGNDASNFANALL---BEVLTAKTWVTAPEAIRGRVQGTWYQKTVQ--LPAGTKYVA 166
 Db 123 ASSPEELTNIIIDKVMYEKATMDKSYLEKVLIIAG--ADYSNNSQVGQPTIKYGMQYYY 180
 Qy 167 FRHFGCTDFFWINLDDVEIKANGKRADFTETTFESSTHGBAPAEWTTIDADGDGQGL--C 224
 Db 181 NQEHGVTDVY-----NYLKAPYTCGYSHLNTGVSFANYT---AHGSETAWADPL 226
 Qy 225 LSSGQLGWLTL-----AHGCTNVVASFWSNGMALNPDNYLISKDVTGATKVKYVAVND 277
 Db 227 LITSQKALTNKDKYFLAIGNCCITAQFDY-----VQP-----CFGEVITRVKEKAYAYIG 278
 Qy 278 GFPGDH-----YAVMISKTGTNAGDFTVVF--EETPNGINKGGARFGLSTEA 322
 Db 279 SSNSYWGEDYYWSVGANAVFGVQPTFEGTSMGSYDATFLEDSVNTVN----- 326
 Qy 323 DGAKPOSVWERTVDLPAQT-----KYVAFRHYNCSDLYNILLDDIQFTMGSGSPPTDYT- 377
 Db 327 -----SIMWAGNLAATHAGNIGNITHIGAHY--WEAYHVLGDSVMPYRAMPKNTNTYTL 379
 Qy 378 -----YTVYRDGKIKKEGLTETTFEEDGVAT-----GNHE 407
 Db 380 PASLPQNASYSIQASAGSVVAISKDGLVYGTGVANAS-----GVATYSMTKQITENGNYD 435
 Qy 408 YCVEVKYTAGVSPKECVNVTVPQFNPNQNLTGSAVGQKVTLLKWDAPN 456
 Db 436 VWITRSNLYPVIKIQVG---EPSYPQPVSNLTATTQGGKVTLLKWEAPS 481

Search completed: May 18, 2004, 11:52:21
 Job time : 36.3211 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:33:39 ; Search time 45.2755 Seconds

(without alignments)

2739.638 Million cell updates/sec

Title: US-08-570-311-22

Perfect score: 2388

Sequence: 1 GTPNPNNPNCPTTLSESF.....PKVCNVNTINPQFNPNQNL 439

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 2388 | 100.0 | 439 | 2 | AAR96024 P. gingiv |
| 2 | 2388 | 100.0 | 439 | 2 | Aaw69492 Haemagglu |
| 3 | 2388 | 100.0 | 2628 | 2 | AAR96030 P. gingiv |
| 4 | 2388 | 100.0 | 2628 | 2 | Aaw69488 Haemagglu |
| 5 | 2351 | 98.5 | 456 | 2 | AAR96023 P. gingiv |
| 6 | 2351 | 98.5 | 456 | 2 | AAR96022 P. gingiv |
| 7 | 2351 | 98.5 | 456 | 2 | Aaw69491 Haemagglu |
| 8 | 2351 | 98.5 | 456 | 2 | AAR96490 Haemagglu |
| 9 | 2321 | 97.2 | 450 | 2 | AAR96021 P. gingiv |
| 10 | 2321 | 97.2 | 450 | 2 | Aaw69489 Haemagglu |
| 11 | 2058 | 86.2 | 1687 | 2 | AAR96033 P. gingiv |
| 12 | 2058 | 86.2 | 1687 | 2 | Aaw69495 Haemagglu |
| 13 | 2058 | 86.2 | 1704 | 2 | AAR70188 Arg-gingi |
| 14 | 2058 | 86.2 | 1704 | 2 | Aaw34843 Arg-gingi |
| 15 | 2058 | 86.2 | 1704 | 3 | AAY67396 Arg-gingi |
| 16 | 2058 | 86.2 | 1704 | 4 | Aau08938 P. gingiv |
| 17 | 2055 | 86.1 | 1087 | 2 | AAR96028 P. gingiv |
| 18 | 2055 | 86.1 | 1087 | 2 | Aaw69486 Haemagglu |
| 19 | 2055 | 86.1 | 1358 | 2 | AAR96032 P. gingiv |
| 20 | 2055 | 86.1 | 1358 | 2 | Aaw69494 Haemagglu |
| 21 | 2048 | 85.8 | 1706 | 2 | Aaw24786 PrtR anti |
| 22 | 2025.5 | 84.8 | 1732 | 2 | AAR96029 P. gingiv |
| 23 | 2025.5 | 84.8 | 1732 | 2 | Aaw24787 PrtK anti |
| 24 | 2025.5 | 84.8 | 1732 | 2 | Aaw69487 Haemagglu |
| 25 | 727 | 30.4 | 135 | 6 | ABP55081 Porphyrom |

ALIGNMENTS

RESULT 1

AAR96024

ID AAR96024 standard; protein; 439 AA.

AC AAR96024;

XX 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)

XX P. gingivalis haemagglutinin hagA HAREP4 product.

DE Haemagglutinin; hagA; periodontal disease; vaccine; antibody; HAREP4.

XX Porphyromonas gingivalis; strain 381.

OS WO9617936-A2.

XX 13-JUN-1996.

PD 11-DEC-1995; 95WO-US016108.

XX 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.

PA (UABR-) UAB RES FOUND.

XX Progulske-Fox A, Tunwasorn S, Lepine G, Han N, Lantz M, Patti JM;

XX WPI; 1996-287181/29.

XX N-PSDB; AAT30648.

XX Porphyromonas gingivalis genes and proteins - used in the detection and

XX vaccination against periodontal disease.

XX Claim 4; Page 114-115; 153pp; English.

XX HAREP4 (AAR96024) is the product of the HAREP4 repeat unit (AAT30648) of

XX the hagA gene (AAR30654) of P. gingivalis 318. It forms part of

XX haemagglutinin hagA (see also AAR96030). HAREP4 and other hagA repeat

XX unit products (see also AAR96021-23) can be obtd. from transformed host

XX cells and used as vaccines to protect humans or animals against

XX periodontal disease. Expression in Salmonella cells allows prodn. of live

XX vaccine. HAREP4 can also be used to detect the presence of anti-P.

XX gingivalis antibodies and to raise monoclonal antibodies for diagnostic

XX appln. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 439 AA;

Query Match 100.0%; Score 2388; DB 2; Length 439;
 Best Local Similarity 100.0%; Pred. No. 1.2e-185;
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTLTSESFENGIPASWKTIDADGNNWTTTPPGTSPAGHNSAICV 60
 DB 1 GTPNPNPNPGTTLTSESFENGIPASWKTIDADGNNWTTTPPGTSPAGHNSAICV 60

QY 61 SSASYINFEQPNPNLYLTPSLPGGGTLTFWVCAQDANYASEHVAVASSTGNDASN 120
 DB 61 SSASYINFEQPNPNLYLTPSLPGGGTLTFWVCAQDANYASEHVAVASSTGNDASN 120

QY 121 FANALLEVLTAKTWVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
 DB 121 FANALLEVLTAKTWVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

QY 181 DEVEIKANGKRAADTFETTESSTHGAPAEWTTIDADGGQGWCLSSGQLDMLTAHGNTN 240
 DB 181 DEVEIKANGKRAADTFETTESSTHGAPAEWTTIDADGGQGWCLSSGQLDMLTAHGNTN 240

QY 241 VVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGFGPDHYAVMISKTGTNAGDFTV 300
 DB 241 VVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGFGPDHYAVMISKTGTNAGDFTV 300

QY 301 VFEETPNGINKGARFGLSTEDGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILL 360
 DB 301 VFEETPNGINKGARFGLSTEDGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILL 360

QY 361 DDIOFTMGSGPTDPTDYTVYVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 DB 361 DDIOFTMGSGPTDPTDYTVYVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420

QY 421 KVCNVNTINPTQFNPVQNL 439
 DB 421 KVCNVNTINPTQFNPVQNL 439

RESULT 2

AAW69492
 ID AAW69492 standard; protein; 439 AA.
 AC AAW69492;
 DT 22-DEC-1998 (first entry)
 DE Haemagglutinin protein hgaA, Harep4.
 XX Haemagglutinin protein; periodontal disease; vaccine; hgaA.
 KW Porphyrromonas gingivalis.
 OS US5824791-A.
 PN 20-OCT-1998.
 XX 11-DEC-1995; 95US-00570311.
 XX 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00353485.
 XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulskie-Fox A, Lepine G;
 DR WPI; 1998-582627/49.
 DR N-PSDB; AAV58879.
 XX Isolated Porphyrromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease poly(peptide(s)).
 XX Claim 1; Col 139-144; 101pp; English.

XX This sequence is encoded by a Porphyrromonas gingivalis gene of the
 CC invention. This sequence represents the hgaA haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease

XX SQ Sequence 439 AA;
 Query Match 100.0%; Score 2388; DB 2; Length 439;
 Best Local Similarity 100.0%; Pred. No. 1.2e-185;
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTLTSESFENGIPASWKTIDADGNNWTTTPPGTSPAGHNSAICV 60
 DB 1 GTPNPNPNPGTTLTSESFENGIPASWKTIDADGNNWTTTPPGTSPAGHNSAICV 60

QY 61 SSASYINFEQPNPNLYLTPSLPGGGTLTFWVCAQDANYASEHVAVASSTGNDASN 120
 DB 61 SSASYINFEQPNPNLYLTPSLPGGGTLTFWVCAQDANYASEHVAVASSTGNDASN 120

QY 121 FANALLEVLTAKTWVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180
 DB 121 FANALLEVLTAKTWVTAPAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINL 180

QY 181 DEVEIKANGKRAADTFETTESSTHGAPAEWTTIDADGGQGWCLSSGQLDMLTAHGNTN 240
 DB 181 DEVEIKANGKRAADTFETTESSTHGAPAEWTTIDADGGQGWCLSSGQLDMLTAHGNTN 240

QY 241 VVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGFGPDHYAVMISKTGTNAGDFTV 300
 DB 241 VVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGFGPDHYAVMISKTGTNAGDFTV 300

QY 301 VFEETPNGINKGARFGLSTEDGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILL 360
 DB 301 VFEETPNGINKGARFGLSTEDGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLNYILL 360

QY 361 DDIOFTMGSGPTDPTDYTVYVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420
 DB 361 DDIOFTMGSGPTDPTDYTVYVRDGTKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSP 420

QY 421 KVCNVNTINPTQFNPVQNL 439
 DB 421 KVCNVNTINPTQFNPVQNL 439

RESULT 3

AAW96030
 ID AAR96030 standard; protein; 2628 AA.
 XX AAR96030;
 XX 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX P. gingivalis haemagglutinin hgaA.
 XX Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.
 KW Porphyrromonas gingivalis; strain 381.
 OS WO9617936-A2.
 XX 13-JUN-1996.
 PD 11-DEC-1995; 95WO-US016108.
 PF 09-DEC-1994; 94US-00353485.

XX (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Proguiske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 XX WPI; 1996-287181/29.
 DR N-PSDB; AAT30654.
 XX
 XX Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 6; Page 93-101; 153pp; English.
 XX
 CC P. gingivalis 381 haemagglutinin hga (AAR96030) was identified as the
 CC product of the hga gene (AAT30654) isolated as an EcoRV fragment of
 CC genomic DNA. The haemagglutinin, or portions of it (see also AAR96021-
 CC 24), can be obtd. from transformed host cells and used as a vaccine to
 CC protect humans or animals against periodontal disease. Expression in
 CC Salmonella cells allows prodn. of a live vaccine. The haemagglutinin can
 CC also be used to detect the presence of anti-P. gingivalis antibodies and
 CC to raise monoclonal antibodies for diagnostic appln. (Updated on 16-OCT-
 CC 2003 to standardise OS field)
 XX
 XX Sequence 2628 AA;
 SQ
 Query Match 100.0%; Score 2388; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 1.5e-184;
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTPNPNPNPQTTLSEFENGIPASWKTIIDADGNNWTTTPPGGTSFAGHNSAICV 60
 DB 1862 GTPNPNPNPQTTLSEFENGIPASWKTIIDADGNNWTTTPPGGTSFAGHNSAICV 1921
 QY 61 SSASVINFEQPNPNYLVTPELSLPGGGTLFWVCAQDANYASEHYAVASSTGNDASN 120
 DB 1922 SSASVINFEQPNPNYLVTPELSLPGGGTLFWVCAQDANYASEHYAVASSTGNDASN 1981
 QY 121 FANALLEEVLTAKTVVTAPPAIRTRVQGTWYQKTQVLPAGTKYVAFRHFGCTDFFWINL 180
 DB 1982 FANALLEEVLTAKTVVTAPPAIRTRVQGTWYQKTQVLPAGTKYVAFRHFGCTDFFWINL 2041
 QY 181 DEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAHGGTN 240
 DB 2042 DEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAHGGTN 2101
 QY 241 VVASFWSNGMALNPDPNLYLSKDVTKATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
 DB 2102 VVASFWSNGMALNPDPNLYLSKDVTKATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 2161
 QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
 DB 2162 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221
 QY 361 DDIOFTMGSSPTDYYTVYRDGKIKKEGLTETTTFFEDGVATGNHCEVCVEKYTAGVSP 420
 DB 2222 DDIOFTMGSSPTDYYTVYRDGKIKKEGLTETTTFFEDGVATGNHCEVCVEKYTAGVSP 2281
 QY 421 KVCNVNTINPTQFNPNQNL 439
 DB 2282 KVCNVNTINPTQFNPNQNL 2300
 RESULT 4
 ID AAW69488
 XX AAW69488 standard; protein; 2628 AA.
 AC AAW69488;
 XX
 XX 22-DEC-1998 (first entry)
 XX Haemagglutinin protein hga.
 DE

XX Haemagglutinin protein; periodontal disease; vaccine; hga.
 OS Porphyromonas gingivalis.
 XX
 PN US824791-A.
 XX
 XX 20-OCT-1998.
 PD
 XX 11-DEC-1995; 95US-00570311.
 PF
 XX 08-SEP-1988; 88US-00241640.
 PR 25-JAN-1991; 91US-00647119.
 PR 09-DEC-1994; 94US-00953485.
 XX
 PA (UYFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 XX Patti JM, Han N, Lantz M, Tumwasorn S, Proguiske-Fox A, Lepine G;
 PI WPI; 1998-582627/49.
 DR N-PSDB; AAV58875.
 XX
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
 PT protease poly:peptide(s)).
 XX
 XX Claim 1; Col 91-110; 101pp; English.
 PS
 XX This sequence is encoded by a Porphyromonas gingivalis gene of the
 CC invention. This sequence represents the hga haemagglutinin protein. The
 CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 SQ Sequence 2628 AA;
 Query Match 100.0%; Score 2388; DB 2; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 1.5e-184;
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTPNPNPNPQTTLSEFENGIPASWKTIIDADGNNWTTTPPGGTSFAGHNSAICV 60
 DB 1862 GTPNPNPNPQTTLSEFENGIPASWKTIIDADGNNWTTTPPGGTSFAGHNSAICV 1921
 QY 61 SSASVINFEQPNPNYLVTPELSLPGGGTLFWVCAQDANYASEHYAVASSTGNDASN 120
 DB 1922 SSASVINFEQPNPNYLVTPELSLPGGGTLFWVCAQDANYASEHYAVASSTGNDASN 1981
 QY 121 FANALLEEVLTAKTVVTAPPAIRTRVQGTWYQKTQVLPAGTKYVAFRHFGCTDFFWINL 180
 DB 1982 FANALLEEVLTAKTVVTAPPAIRTRVQGTWYQKTQVLPAGTKYVAFRHFGCTDFFWINL 2041
 QY 181 DEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAHGGTN 240
 DB 2042 DEVEIKANGKRAADFTETPESSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAHGGTN 2101
 QY 241 VVASFWSNGMALNPDPNLYLSKDVTKATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
 DB 2102 VVASFWSNGMALNPDPNLYLSKDVTKATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 2161
 QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
 DB 2162 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILL 2221
 QY 361 DDIOFTMGSSPTDYYTVYRDGKIKKEGLTETTTFFEDGVATGNHCEVCVEKYTAGVSP 420
 DB 2222 DDIOFTMGSSPTDYYTVYRDGKIKKEGLTETTTFFEDGVATGNHCEVCVEKYTAGVSP 2281
 QY 421 KVCNVNTINPTQFNPNQNL 439
 DB 2282 KVCNVNTINPTQFNPNQNL 2300

```

RESULT 5
AAR96023
ID AAR96023 standard; protein; 456 AA.
XX
AC AAR96023;
XX
DT 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
DE P. gingivalis haemagglutinin hAgA Harep3 product.
XX
KW Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harep3.
XX
OS Porphyromonas gingivalis; strain 381.
XX
FN WO9617936-A2.
XX
PD 13-JUN-1996.
XX
PF 11-DEC-1995; 95WO-US016108.
XX
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Progukske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR WPI; 1996-287181/29.
DR N-PSDB; AAT30647.
XX
PT Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
XX
PS Claim 4; Page 110-113; 153pp; English.
XX
CC Harep3 (AAR96023) is the product of the Harep3 repeat unit (AAT30647) of
CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
CC haemagglutinin hAgA (see also AAR96030). Harep3 and other hAgA repeat
CC unit products (see also AAR96021-22 and AAR96024) can be obt'd. from
CC transformed host cells and used as vaccines to protect humans or animals
CC against periodontal disease. Expression in Salmonella cells allows prodn.
CC of live vaccine. Harep1-4 can also be used to detect the presence of anti
CC -P. gingivalis antibodies and to raise monoclonal antibodies for
CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 456 AA;

Query Match 98.5%; Score 2351; DB 2; Length 456;
Best Local Similarity 98.2%; Pred. No. 1.3e-182;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTLTSSFENGIPASWKTIDADGNGNWTTPPPGGTSFAGHNSAICV 60
DB 1 GTPNPNPNPGTTLTSSFENGIPASWKTIDADGNGNWTTPPPGGTSFAGHNSAICA 60
QY 61 SSASYNIFEGPQNPNDYLTPELSLPGGTLTFWVCAQDANYASEHYAVASSTGNDASN 120
DB 61 SSASYNIFEGPQNPNDYLTPELSLPGGTLTFWVCAQDANYASEHYAVASSTGNDASN 120
QY 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 180
DB 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 180
QY 181 DEVEIKANGKRAPTTFESTTGEPAEWTTTIDADGGQGWCLSSGQLDWLTAGGTN 240
DB 181 DVEIKANGKRAPTTFESTTGEPAEWTTTIDADGGQGWCLSSGQLGWLTAGGTN 240
QY 241 VVASFNSNGMALPNPNVLISSKDVGTATKVKYYAVANDGPFCDHYAVWISKTGNAGDFTV 300
DB 241 VVASFNSNGMALPNPNVLISSKDVGTATKVKYYAVANDGPFCDHYAVWISKTGNAGDFTV 300
QY 301 VFEBETENGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILL 360

```

```

DB 301 VFEBETENGKKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILL 360
QY 361 DDIOFTMGSGSPTTDYTYVYRDGTKIKEGLTETTTEDDGVATGNHEYCVVEVKYTAGVSP 420
DB 361 DDIOFTMGSGSPTTDYTYVYRDGTKIKEGLTETTTEDDGVATGNHEYCVVEVKYTAGVSP 420
QY 421 KVCVNVTVINPTQFNPVQNL 439
DB 421 KECVNVTVDPQFNPVQNL 439

RESULT 6
AAR96022
ID AAR96022 standard; protein; 456 AA.
XX
AC AAR96022;
XX
DT 16-OCT-2003 (revised)
DT 04-SEP-1996 (first entry)
XX
DE P. gingivalis haemagglutinin hAgA Harep2 product.
XX
KW Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harep2.
XX
OS Porphyromonas gingivalis; strain 381.
XX
FN WO9617936-A2.
XX
PD 13-JUN-1996.
XX
PF 11-DEC-1995; 95WO-US016108.
XX
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL ) UNIV FLORIDA.
PA (UABR-) UAB RES FOUND.
XX
PI Progukske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX
DR WPI; 1996-287181/29.
DR N-PSDB; AAT30646.
XX
PT Porphyromonas gingivalis genes and proteins - used in the detection and
PT vaccination against periodontal disease.
XX
PS Claim 4; Page 107-108; 153pp; English.
XX
CC Harep2 (AAR96022) is the product of the Harep2 repeat unit (AAT30646) of
CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
CC haemagglutinin hAgA (see also AAR96030). Harep2 and other hAgA repeat
CC unit products (see also AAR96021 and AAR96023-24) can be obt'd. from
CC transformed host cells and used as vaccines to protect humans or animals
CC against periodontal disease. Expression in Salmonella cells allows prodn.
CC of live vaccine. Harep1-4 can also be used to detect the presence of anti
CC -P. gingivalis antibodies and to raise monoclonal antibodies for
CC diagnostic appln. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ Sequence 456 AA;

Query Match 98.5%; Score 2351; DB 2; Length 456;
Best Local Similarity 98.2%; Pred. No. 1.3e-182;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTPNPNPNPGTTLTSSFENGIPASWKTIDADGNGNWTTPPPGGTSFAGHNSAICV 60
DB 1 GTPNPNPNPGTTLTSSFENGIPASWKTIDADGNGNWTTPPPGGTSFAGHNSAICA 60
QY 61 SSASYNIFEGPQNPNDYLTPELSLPGGTLTFWVCAQDANYASEHYAVASSTGNDASN 120
DB 61 SSASYNIFEGPQNPNDYLTPELSLPGGTLTFWVCAQDANYASEHYAVASSTGNDASN 120
QY 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFCTDFFWINL 180

```

```

Db 121 FANALLEVLTAKTVTVAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWNL 180
QY 181 DEVEIKANGKRAADFTETFESSTHGEAPAEWTTIDADGGQWLCSSGQDGLWLTAGGNT 240
Db 181 DVEIKANGKRAADFTETFESSTHGEAPAEWTTIDADGGQWLCSSGQDGLWLTAGGNT 240
QY 241 VVASFSWNGMALNPNDYLSKDVGTGATKVKYKYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
Db 241 VVASFSWNGMALNPNDYLSKDVGTGATKVKYKYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILL 360
Db 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILL 360
QY 361 DDIQFTMGSSPTDVTYTVYRDGTIKKGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
Db 361 DDIQFTMGSSPTDVTYTVYRDGTIKKGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
QY 421 KVCNVNTINPTQFNPNQNL 439
Db 421 KECVNTVDPVQFNPNQNL 439

RESULT 7
AAW69491
ID AAW69491 standard; protein; 456 AA.
XX
AC AAW69491;
XX
DT 22-DEC-1998 (first entry)
DE Haemagglutinin protein hAgA, HAre3.
XX Haemagglutinin protein; periodontal disease; vaccine; hAgA.
XX Porphyromonas gingivalis.
XX US5824791-A.
XX 20-OCT-1998.
XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58878.
XX
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
protease poly:peptide(s)).
XX Claim 1; Col 133-138; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
invention. This sequence represents the hAgA haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 456 AA;

```

Query Match 98.5%; Score 2351; DB 2; Length 456;
 Best Local Similarity 98.2%; Pred. No. 1.3e-182;
 Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 GTPENPNPNPGTTLTSSSFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAICV 60
Db 1 GTPENPNPNPGTTLTSSSFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAICV 60
QY 61 SSASINIFEGPQNPONLYLTPELSIPGGGTLTFWVCAQDANYASBHYAVYASSTGNDASN 120
Db 61 SSASINIFEGPQNPONLYLTPELSIPGGGTLTFWVCAQDANYASBHYAVYASSTGNDASN 120
QY 121 FANALLEVLTAKTVTVAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWNL 180
Db 121 FANALLEVLTAKTVTVAPEAIRGRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWNL 180
QY 181 DEVEIKANGKRAADFTETFESSTHGEAPAEWTTIDADGGQWLCSSGQDGLWLTAGGNT 240
Db 181 DVEIKANGKRAADFTETFESSTHGEAPAEWTTIDADGGQWLCSSGQDGLWLTAGGNT 240
QY 241 VVASFSWNGMALNPNDYLSKDVGTGATKVKYKYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
Db 241 VVASFSWNGMALNPNDYLSKDVGTGATKVKYKYAVNDGPPGDHYAVMISKTGTNAGDFTV 300
QY 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILL 360
Db 301 VFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHNCSDLNYILL 360
QY 361 DDIQFTMGSSPTDVTYTVYRDGTIKKGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
Db 361 DDIQFTMGSSPTDVTYTVYRDGTIKKGLTETTFEEDGVATGNHEYCVVKYTAGVSP 420
QY 421 KVCNVNTINPTQFNPNQNL 439
Db 421 KECVNTVDPVQFNPNQNL 439

RESULT 8
AAW69490
ID AAW69490 standard; protein; 456 AA.
XX
AC AAW69490;
XX
DT 22-DEC-1998 (first entry)
DE Haemagglutinin protein hAgA, HAre2.
XX Haemagglutinin protein; periodontal disease; vaccine; hAgA.
XX Porphyromonas gingivalis.
XX US5824791-A.
XX 20-OCT-1998.
XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYFL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulske-Fox A, Lepine G;
XX WPI; 1998-582627/49.
DR N-PSDB; AAV58877.
XX
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
protease poly:peptide(s)).
XX Claim 1; Col 127-132; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
invention. This sequence represents the hAgA haemagglutinin protein. The

```

CC polypeptides are used to produce antibodies to organisms associated with
 CC periodontal disease. The antibodies are also used in purification and
 CC identification procedures. The genes and polypeptides are used as
 CC vaccines against periodontal disease
 XX
 SQ Sequence 456 AA;

Query Match 98.5%; Score 2351; DB 2; Length 456;
 Best Local Similarity 98.2%; Pred. No. 1.3e-182;
 Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 GTPNPNPNPGTTLTSSSEFNGIPASWKTIDADGDNNTTTPPPGGTSFAGHNSAICV 60
 Db 1 GTPNPNPNPGTTLTSSSEFNGIPASWKTIDADGDNNTTTPPPGGTSFAGHNSAICA 60
 Qy 61 SSASVINFEQPNPNYLVTPSLPGGTLTFWCAQDANYASEHYAVYASSTGNDASN 120
 Db 61 SSASVINFEQPNPNYLVTPSLPGGTLTFWCAQDANYASEHYAVYASSTGNDASN 120
 Qy 121 FANALLEVLTAKTVVTAPAIRGTRVQGTWYQKTVQLPAGTKYVAFRRHFGCTDFFWNL 180
 Db 121 FANALLEVLTAKTVVTAPAIRGTRVQGTWYQKTVQLPAGTKYVAFRRHFGCTDFFWNL 180
 Qy 181 DEVEIKANGKADFTETTESSTHGEAPAEWTTIDADGCGQWCLSSGQLWLTAGGNTN 240
 Db 181 DDVEIKANGKADFTETTESSTHGEAPAEWTTIDADGCGQWCLSSGQLWLTAGGNTN 240
 Qy 241 VVAFSPWNGMALPNPNYILISKDVTGATKVKYYAVNDGFGPDHVAVMISKGTNAGDFTV 300
 Db 241 VVAFSPWNGMALPNPNYILISKDVTGATKVKYYAVNDGFGPDHVAVMISKGTNAGDFTV 300
 Qy 301 VFETPNKNGKARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRRHNCSDLYILL 360
 Db 301 VFETPNKNGKARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRRHNCSDLYILL 360
 Qy 361 DDIOFTMGSGPTDYYTVVYRDGKIKRGLTETTFEEDGVATGNHCVKVTAGVSP 420
 Db 361 DDIOFTMGSGPTDYYTVVYRDGKIKRGLTETTFEEDGVATGNHCVKVTAGVSP 420
 Qy 421 KVCNVNTINPTQFNPQNVL 439
 Db 421 KECNVNTVDPVQFNPQNVL 439

RESULT 9
 AAR96021
 ID AAR96021 standard; protein; 450 AA.
 XX
 AC AAR96021;
 XX
 DT 16-OCT-2003 (revised)
 DT 04-SEP-1996 (first entry)
 XX
 DE P. gingivalis haemagglutinin hAgA Harepl product.
 XX
 KW Haemagglutinin; hAgA; periodontal disease; vaccine; antibody; Harepl.
 XX
 OS Porphyromonas gingivalis; strain 381.
 XX
 PN W09617936-A2.
 XX
 PD 13-JUN-1996.
 XX
 PF 11-DEC-1995; 95WO-US016108.
 XX
 PR 09-DEC-1994; 94US-00353485.
 XX
 PA (UFL) UNIV FLORIDA.
 PA (UABR-) UAB RES FOUND.
 XX
 PI Progulske-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
 DR WPI; 1996-287181/29.

DR N-PSDB; AAT30645.
 XX
 PT Porphyromonas gingivalis genes and proteins - used in the detection and
 PT vaccination against periodontal disease.
 XX
 XX Claim 4; Page 103-104; 153pp; English.
 XX
 CC Harepl (AAR96021) is the product of the Harepl repeat unit (AAT30645) of
 CC the hAgA gene (AAT30654) of P. gingivalis 318. It forms part of
 CC haemagglutinin hAgA (see also AAR96030). Harepl and other hAgA repeat
 CC unit products (see also AAR96022-24) can be obtd. from transformed host
 CC cells and used as vaccines to protect humans or animals against
 CC periodontal disease. Expression in Salmonella cells allows prodn. of live
 CC vaccine. Harepl-4 can also be used to detect the presence of anti-P.
 CC gingivalis antibodies and to raise monoclonal antibodies for diagnostic
 CC appln. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 450 AA;

Query Match 97.2%; Score 2321; DB 2; Length 450;
 Best Local Similarity 98.4%; Pred. No. 3.4e-180;
 Matches 426; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Qy 7 PNPNGPTTLTSSSEFNGIPASWKTIDADGDNNTTTPPPGGTSFAGHNSAICVSSASYI 66
 Db 1 PNPNGPTTLTSSSEFNGIPASWKTIDADGDNNTTTPPPGGTSFAGHNSAICASSASYI 60
 Qy 67 NFEQPNPNYLVTPSLPGGTLTFWCAQDANYASEHYAVYASSTGNDASNANALL 126
 Db 61 NFEQPNPNYLVTPSLPGGTLTFWCAQDANYASEHYAVYASSTGNDASNANALL 120
 Qy 127 EVLTAKTVVTAPAIRGTRVQGTWYQKTVQLPAGTKYVAFRRHFGCTDFFWNLDEVEIK 186
 Db 121 EVLTAKTVVTAPAIRGTRVQGTWYQKTVQLPAGTKYVAFRRHFGCTDFFWNLDDVEIK 180
 Qy 187 ANGKADFTETTESSTHGEAPAEWTTIDADGCGQWCLSSGQLWLTAGGNTNVASFS 246
 Db 181 ANGKADFTETTESSTHGEAPAEWTTIDADGCGQWCLSSGQLWLTAGGNTNVASFS 240
 Qy 247 WNGMALPNPNYILISKDVTGATKVKYYAVNDGFGPDHVAVMISKGTNAGDFTVVFETP 306
 Db 241 WNGMALPNPNYILISKDVTGATKVKYYAVNDGFGPDHVAVMISKGTNAGDFTVVFETP 300
 Qy 307 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRRHNCSDLYILLDDIOFT 366
 Db 301 NGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRRHNCSDLYILLDDIOFT 360
 Qy 367 MGSPTPTDYYTVVYRDGKIKRGLTETTFEEDGVATGNHCVKVTAGVSPKVCNV 426
 Db 361 MGSPTPTDYYTVVYRDGKIKRGLTETTFEEDGVATGNHCVKVTAGVSPKVCNV 420
 Qy 427 TINPTQFNPQNVL 439
 Db 421 TVDPVQFNPQNVL 433

RESULT 10
 AAW69489
 ID AAW69489 standard; protein; 450 AA.
 XX
 AC AAW69489;
 XX
 DT 22-DEC-1998 (first entry)
 DE Haemagglutinin protein hAgA, Harepl.
 KW Haemagglutinin protein; periodontal disease; vaccine; hAgA.
 OS Porphyromonas gingivalis.
 PN US5824791-A.
 XX
 PD 20-OCT-1998.


```

XX 11-DEC-1995; 95US-00570311.
XX 08-SEP-1988; 88US-00241640.
XX 25-JAN-1991; 91US-00647119.
XX 09-DEC-1994; 94US-00353485.
XX (UYEL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX Patti JM, Han N, Lantz M, Tumwasorn S, Progulsk-Fox A, Lepine G;
XX WPI; 1998-582627/49.
XX N-PSDB; AAV58876.
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
XX protease poly.peptide(s).
XX Claim 1; Col 121-126; 101pp; English.
XX This sequence is encoded by a Porphyromonas gingivalis gene of the
XX invention. This sequence represents the hAgA haemagglutinin protein. The
XX polypeptides are used to produce antibodies to organisms associated with
XX periodontal disease. The antibodies are also used in purification and
XX identification procedures. The genes and polypeptides are used as
XX vaccines against periodontal disease
XX Sequence 450 AA;

Query Match 97.2%; Score 2321; DB 2; Length 450;
Best Local Similarity 98.4%; Pred. No. 3.4e-180;
Matches 426; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 PNPNTTLLSSEFENGIPASWKTIADGNGNNTTTPPGGTSFAGHNSAICVSSASYI 66
Db 1 PNPNTTLLSSEFENGIPASWKTIADGNGNNTTTPPGGTSFAGHNSAICVSSASYI 60
QY 67 NFEQPNDNYLVTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 126
Db 61 NFEQPNDNYLVTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASNFANALL 120
QY 127 ESVLTAKTVTAPAEIRGTVOGTWYOKTVOLPAGTKYVAFRHFCTDFWINLDEVEIK 186
Db 121 ESVLTAKTVTAPAEIRGTVOGTWYOKTVOLPAGTKYVAFRHFCTDFWINLDEVEIK 180
QY 187 ANGKGRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGGNTNVVASF 246
Db 181 ANGKGRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGGNTNVVASF 240
QY 247 WNGMALNPNDYLSKDVGTATKYYAVNDGPPGDHYAVMISKTGTNAGDFTVPEETP 306
Db 241 WNGMALNPNDYLSKDVGTATKYYAVNDGPPGDHYAVMISKTGTNAGDFTVPEETP 300
QY 307 NGINKGARGFLSTEADGAKPQSVLERTVDLPAGTKYVAFRHYNCSDLNILLDDIQT 366
Db 301 NGINKGARGFLSTEADGAKPQSVLERTVDLPAGTKYVAFRHYNCSDLNILLDDIQT 360
QY 367 MGSPPTPTDYTYTVYRDGTKKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKCVNV 426
Db 361 MGSPPTPTDYTYTVYRDGTKKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKCVNV 420
QY 427 TNPQFNPNQNL 439
Db 421 TVDPQFNPNQNL 433

RESULT 11
AAR96033
ID AAR96033 standard; protein; 1687 AA.
XX
AC AAR96033;
XX
DT 16-OCT-2003 (revised)

DT 04-SEP-1996 (first entry)
XX P. gingivalis haemagglutinin hAgE.
XX Haemagglutinin; hAgE; periodontal disease; vaccine; antibody.
XX Porphyromonas gingivalis; strain FDC381.
XX WO9617936-A2.
XX 13-JUN-1996.
XX 11-DEC-1995; 95WO-US016108.
XX 09-DEC-1994; 94US-00353485.
XX (UYEL ) UNIV FLORIDA.
XX (UABR-) UAB RES FOUND.
XX Progulsk-Fox A, Tumwasorn S, Lepine G, Han N, Lantz M, Patti JM;
XX WPI; 1996-287181/29.
XX N-PSDB; AAT30656.
XX Porphyromonas gingivalis genes and proteins - used in the detection and
XX vaccination against periodontal disease.
XX Claim 5; Page 138-143; 153pp; English.
XX P. gingivalis 381 haemagglutinin hAgE (AAR96033) was identified as the
XX product of a gene (AAT30656) identified in P. gingivalis 318 genomic DNA.
XX The haemagglutinin can be obtd. from transformed host cells and used as a
XX vaccine to protect humans or animals against periodontal disease.
XX Expression in Salmonella cells allows prodn. of a live vaccine. The
XX haemagglutinin can also be used to detect the presence of anti-P.
XX gingivalis antibodies and to raise monoclonal antibodies for diagnostic
XX appln. (Updated on 16-OCT-2003 to standardise OS field)
XX Sequence 1687 AA;

Query Match 86.2%; Score 2058; DB 2; Length 1687;
Best Local Similarity 88.8%; Pred. No. 5.8e-158;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

QY 3 PNPENPNFGTTLSESFENGIPASWKTIADGNGNNTTTPPGGTSFAGHNSAICVSS 62
Db 935 PNPENPNFGTTLSESFENGIPASWKTIADGNGNNTTTPPGGTSFAGHNSAICVSS 991
QY 63 ASY-INFEQPNDNYLVTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 992 ESEGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051
QY 122 ANALLEEVLTAKTVTAPAEIRGTVOGTWYOKTVOLPAGTKYVAFRHFCTDFWINL 181
Db 1052 TWALLEEITAGK-VRSPPAIRG-RIOGTWRQKTVDLPAGTKYVAFRHFQSTDMFTID 1109
QY 182 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGGNTV 241
Db 1110 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDLWLTAGGNTV 1169
QY 242 VASFSWNGMALNPNDYLSKDVGTATKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 301
Db 1170 VASFSWNGMALNPNDYLSKDVGTATKYYAVNDGPPGDHYAVMISKTGTNAGDFTV 1229
QY 302 FEETPNGINKGARGFLSTEADGAKPQSVLERTVDLPAGTKYVAFRHYNCSDLNILL 361
Db 1230 FEETPNGINKGARGFLSTEADGAKPQSVLERTVDLPAGTKYVAFRHYNCSDLNILL 1289
QY 362 DIQFTMGSPPTDYTYTVYRDGTKKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
Db 1290 DIQFTMGSPPTDYTYTVYRDGTKKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1349
QY 422 VCUNVTINPTQFNPNQNL 439

```

Db 1350 ECVNVTINPTQFNPKNL 1367
|||||:|
RESULT 12
AAW69495
ID AAW69495 standard; protein; 1687 AA.
XX
AC AAW69495;
XX
DT 22-DEC-1998 (first entry)
XX
DE Haemagglutinin protein hAgE.
XX
KW Haemagglutinin protein; periodontal disease; vaccine; hAgE.
XX
OS Porphyromonas gingivalis.
XX
PN US5824791-A.
XX
PD 20-OCT-1998.
XX
PF 11-DEC-1995; 95US-00570311.
XX
PR 08-SEP-1988; 88US-00241640.
PR 25-JAN-1991; 91US-00647119.
PR 09-DEC-1994; 94US-00353485.
XX
PA (UYFL) UNIV FLORIDA.
PA (UABR) UAB RES FOUND.
XX
PI Patti JM, Han N, Lantz M, Tumwasorn S, Progulsk-Fox A, Lepine G;
XX
DR WPI; 1998-582627/49.
DR N-PSDB; AAV58881.
XX
XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin and/or
PT protease poly:peptide(s).
XX
PS Claim 1; Col 167-182; 101pp; English.
XX
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hAgE haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease
XX
SQ Sequence 1687 AA;
Query Match 86.2%; Score 2058; DB 2; Length 1687;
Best Local Similarity 88.8%; Pred. NO. 5.8e-158;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;
3 PNPENPNPGTTLSESPENGIPASWKTIDADGCGNWTTPPGGTSFAGHNSAICVSS 62
935 PNPENPNPGTTLSESPENGIPASWKTIDADGCGHGWKPGNAPG---IAGYNSGCVYS 991
63 ASY-INPEGPQNPONLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
992 ESFGGGIGVLTDPNLTLPALDLENGGKLTFWCAQDANYASEHYAVYASSTGNDASNF 1051
122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWQKTVQVLPAGTKYVAFRHFGCTDFFWNL 181
1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFGQSTDMFYIDL 1109
182 EVELKANKRADFTETPESSTHGEPAEWTTIDADGCGWLCISSLGOLDLWTAHGGTNV 241
1110 EVELKANKRADFTETPESSTHGEPAEWTTIDADGCGWLCISSLGOLDLWTAHGGTNV 1169
242 VASFSWNGMALNPONYLISKDVTGATKVKYVAVNDGFGPDHYAVMISKTGTNAGDFTVV 301
1170 VASFSWNGMALNPONYLISKDVTGATKVKYVAVNDGFGPDHYAVMISKTGTNAGDFTVV 1229

QY 302 FEETPNKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNVL 361
|||:|
Db 1230 FEETPNKGGARFGLSTEANGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNVL 1289
|||:|
QY 362 DIOFTMGGSPTPDYTYTVYRDGTGKIKEGLTETTFEEDGVATONHVEYCVVEKVTAGVSPK 421
|||:|
Db 1290 DIOFTMGGSPTPDYTYTVYRDGTGKIKEGLTETTFEEDGVATONHVEYCVVEKVTAGVSPK 1349
|||:|
QY 422 VCVNVTINPTQFNPKNL 439
|||:|
Db 1350 ECVNVTINPTQFNPKNL 1367
|||:|
RESULT 13
AAR70188
ID AAR70188 standard; protein; 1704 AA.
XX
AC AAR70188;
XX
DT 25-MAR-2003 (revised)
DT 21-SEP-1995 (first entry)
XX
DE Arg-gingipain-2 prepolyprotein.
XX
KW Arg-gingipain-2; gingivalis; periodontal disease; vaccine;
KW arginine-specific protease.
XX
OS Porphyromonas gingivalis.
XX
FH Key Location/Qualifiers
FT Protein 228..719
FT /label= Protease
FT /note= "corresponds to Arg-gingipain-1"
FT Region 720..1091
FT /label= Hemagglutinin
FT Region 1092..1429
FT /label= Hemagglutinin
FT Region 1430..1704
FT /label= Hemagglutinin
XX
PN W09507286-A1.
XX
PD 16-MAR-1995.
XX
PF 09-SEP-1994; 94WO-US010283.
XX
PR 10-SEP-1993; 93US-00119361.
PR 21-OCT-1993; 93US-00141324.
PR 24-JUN-1994; 94US-00265441.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa J, Barr PJ, Pavloff N;
XX
DR WPI; 1995-123373/16.
DR N-PSDB; AAQ83489.
XX
PT DNA encoding Arg-gingipain proteins - used to develop prods. for
PT detection, treatment and prevention of periodontal disease.
XX
PS Disclosure; Page 70-77; 89pp; English.
XX
CC A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt. AG (AG-
CC 2) were isolated from P. gingivalis strains H66 (ATCC 33277) and W50
CC (ATCC 53973). The sequences of the proteins were used to design PCR
CC primers and probes to isolate AG DNA. Lambda DASH and lambda ZAP
CC libraries were screened with a probe based on amino acids 11-22 of the AG
CC protein to obtain DNA encoding AG-1 (AAQ83484) and AG-2 (AAQ83489). AG-2
CC is a prepolyprotein incorporating AG-1. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 1704 AA;

Query Match 86.2%; Score 2058; DB 2; Length 1704;
 Best Local Similarity 88.8%; Pred. No. 5.9e-158;
 Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTIADGDGNWTTTPPGGTSGFAGHNSAICVSS 62
 DB 952 PNPENPNPGTTLTSESFENGIPASWKTIADGDGHGKPGNAPG---IAGYNSNGCVYS 1008

QY 63 ASY-INFPGPQNPONLYITPDLNGKLTFWVCAQADANYASEHYAVASSTGNDASNF 121
 DB 1009 ESFGLGGIGVLTDPNLYITPALDNGKLTFWVCAQADANYASEHYAVASSTGNDASNF 1068

QY 122 ANALLEEVLTAKTVTAPEAIRGTVQGTWYQKTVDLPAGTKYVAFRHYNCSDLYIILLD 181
 DB 1069 TNALLEETITAG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHYNCSDLYIILLD 1126

QY 182 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDMLTAHGGTNV 241
 DB 1127 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDMLTAHGGTNV 1186

QY 242 VASFSWNGMALNPONLYISKDVTGATKUYVAVNDGFGPDHYAVMISKTGTNAGDFTV 301
 DB 1187 VASFSWNGMALNPONLYISKDVTGATKUYVAVNDGFGPDHYAVMISKTGTNAGDFTV 1246

QY 302 FEETPNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIILLD 361
 DB 1247 FEETPNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIILLD 1306

QY 362 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
 DB 1307 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1366

QY 422 VCVNVTINPTQNPVQNL 439
 DB 1367 ECVNVTINPTQNPVQNL 1384

RESULT 14
 AAW34843
 ID AAW34843 standard; protein; 1704 AA.
 XX AC AAW34843;
 XX XX
 DT 03-JUN-1998 (first entry)
 DE Arg-gingipain high molecular weight prepolyprotein sequence.
 KW Arg-specific gingipain protease; gingivalis; periodontal disease;
 KW vaccine; infection.
 XX OS Porphyromonas gingivalis.
 XX XX
 FH Key Location/Qualifiers
 FT Protein 1..227
 FT /note= "precursor protein"
 XX
 FN WO9734629-A1.
 XX
 PD 25-SEP-1997.
 XX
 PF 21-MAR-1997; 97WO-US004635.
 XX
 PR 22-MAR-1996; 96US-0013945P.
 XX
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA (MORE-) MOREHOUSE SCHOOL MEDICINE.
 XX
 PI Potempa J, Travis J, Genco C;
 XX WPI; 1997-479993/44.
 DR N-PSDB; AAT93872.
 XX

Porphyromonas gingivalis Arg-specific gingipain protease peptide(s) -
 useful for protecting animals and humans from gingivalis and periodontal
 diseases.
 Disclousure; Page 68-73; 95pp; English.
 The present sequence represents an arginine-specific protease of
 Porphyromonas gingivalis. The following peptides, derived from Arg- and
 Lys-specific high molecular weight proteases, offer protection against
 infection: YTVTVYRDGK IKEGLTATTE DGVATGNHE YCVVEKYTAGS VSPKVC (I);
 YTPVEEKONG RMIVIVAKKY (II); QLPFIPDVAC VNGDFLSMP CFARALBRAQ (III);
 GEPNYPQVS NLRATIQGK VTLKWDASTK (IV); GNHEICVEVK YTAGVSEKVC KDVTY (V);
 RMENYEPGR YTEVEEKONG (VI); TFAGFEDTYK RMFMYEPR (VII); DYTIVYRDG
 CC TKIKEGLTAT TEEDGVATGNHE YCVVEKYTAGS VSPKVC (VIII); YTVTVYRDG KIKEGLTATTF
 CC EEDG (IX); RDGKIKEGL TATTEEDGV ATGN (X); KIKEGLTATT FEEDGVATGN HEY (XI)
 CC : KWDAPNGTPN PNPEN PGNPN PGTTLSE (XII); and YTPVEEKENG RMIVIVAKKY
 CC (XIII). They are used in vaccines to protect animals, including humans,
 CC from gingivitis and/or periodontal diseases
 XX
 SQ Sequence 1704 AA;

Query Match 86.2%; Score 2058; DB 2; Length 1704;
 Best Local Similarity 88.8%; Pred. No. 5.9e-158;
 Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKTIADGDGNWTTTPPGGTSGFAGHNSAICVSS 62
 DB 952 PNPENPNPGTTLTSESFENGIPASWKTIADGDGHGKPGNAPG---IAGYNSNGCVYS 1008

QY 63 ASY-INFPGPQNPONLYITPDLNGKLTFWVCAQADANYASEHYAVASSTGNDASNF 121
 DB 1009 ESFGLGGIGVLTDPNLYITPALDNGKLTFWVCAQADANYASEHYAVASSTGNDASNF 1068

QY 122 ANALLEEVLTAKTVTAPEAIRGTVQGTWYQKTVDLPAGTKYVAFRHYNCSDLYIILLD 181
 DB 1069 TNALLEETITAG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHYNCSDLYIILLD 1126

QY 182 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDMLTAHGGTNV 241
 DB 1127 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDMLTAHGGTNV 1186

QY 242 VASFSWNGMALNPONLYISKDVTGATKUYVAVNDGFGPDHYAVMISKTGTNAGDFTV 301
 DB 1187 VASFSWNGMALNPONLYISKDVTGATKUYVAVNDGFGPDHYAVMISKTGTNAGDFTV 1246

QY 302 FEETPNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIILLD 361
 DB 1247 FEETPNGKKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYIILLD 1306

QY 362 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
 DB 1307 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1366

QY 422 VCVNVTINPTQNPVQNL 439
 DB 1367 ECVNVTINPTQNPVQNL 1384

RESULT 15
 AAW67396
 ID AAW67396 standard; protein; 1704 AA.
 XX AC AAW67396;
 XX XX
 DT 25-APR-2000 (first entry)
 DE Arg-gingipain-2 amino acid sequence.
 XX
 KW Arginine specific proteinase; Arg-gingipain; gingipain-2; haemagglutinin;
 KW immunogenic component; vaccine; inflammatory response; tissue damage;
 KW periodontal disease.
 XX
 OS Porphyromonas gingivalis.

```
XX FH Key Location/Qualifiers
XX FT Region 229..719
XX FT /note= "Amino acids 229-719 are specifically claimed"
XX FT Region 720..1185
XX FT /note= "Amino acids 720-1185 are specifically claimed"
XX FT
XX FT
XX PN US6017532-A.
XX PD 25-JAN-2000.
XX XX
XX PF 08-NOV-1994; 94US-00336308.
XX XX
XX PR 10-SEP-1993; 93US-00119361.
XX PR 24-JUN-1994; 94US-00265441.
XX XX
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX PI Potempa JS, Travis J;
XX XX
XX DR WPI; 2000-136659/12.
XX DR N-PSDB; AAZ60181.
XX XX
XX PT New Porphyromonas gingivalis arginine-specific protease preparation
XX PT useful for preparing vaccines against periodontal disease and for
XX PT screening for Arg-gingipain inhibitors.
XX PS Claim 1; Col 29-42; 55pp; English.
XX XX
XX CC This sequence represents a Porphyromonas gingivalis arginine-specific
XX CC proteinase known as Arg-gingipain/gingipain-2 amino acid sequence.
XX CC Gingipain-2 consists of a 50kD protease component non-covalently
XX CC associated with a 44kD haemagglutinin component. The proteinase is
XX CC stimulated by glycine containing peptides and glycine analogues. It is
XX CC inhibited by cysteine protease group specific inhibitors. The protease
XX CC preparation can be used in immunogenic compositions and vaccines against
XX CC inflammatory response and tissue damage caused by P. gingivalis in
XX CC periodontal disease. It can also be used to screen for agents that
XX CC modulate Arg-gingipain proteinase activity inhibitors
XX SQ Sequence 1704 AA;

Query Match 86.2%; Score 2058; DB 3; Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.9e-158;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

Qy 3 PNPNNPNNPGTTLSESPENGIPASWKTIADGDGNNWTTTPPGGTSFAGHNSAICVSS 62
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
952 PNPNNPNNPGTTLSESPENGIPASWKTIADGDGHWKPGNAPG--IAGYNSGCVYS 1008
Qy 63 ASY-INFEQPONPNYLVTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1009 ESFGIGGIGVLTPDNYLITPDLFNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1068
Qy 122 ANALLEVLTAKVTYTAPEARTVQCTWYKTVOLPAGTKYVAFRHFQCTDFEWINLD 181
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1069 TNALLEETITKAG-VRSPEARIG-RIQGTWRQKTVLLPAGTKYVAFRHFQSTDMFYIDL 1126
Qy 182 EVEIKANGKRADFTTFESSTHGEAPAEWTIIDADGGQWMLCLSSGQLDMLTAHGNTV 241
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1127 EVEIKANGKRADFTTFESSTHGEAPAEWTIIDADGGQWMLCLSSGQLDMLTAHGNTV 1186
Qy 242 VASFSWNGMALNPDNYLISKVGTATKVKYKYAVNDGPFQDHYAVMISKTGTNAGDFTV 301
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1187 VASFSWNGMALNPDNYLISKVGTATKVKYKYAVNDGPFQDHYAVMISKTGTNAGDFTV 1246
Qy 302 FEETENGKKGARGLSTEADGAKPQSWLERTVDLPAGTKYVAFRHWNGSDLNLYILLD 361
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1247 FEETENGKKGARGLSTEANGAKPQSWLERTVDLPAGTKYVAFRHWNGSDLNLYILLD 1306
Qy 362 DIQFTMGGSPTPTDYTYVRDGTIKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSPK 421
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 DIQFTMGGSPTPTDYTYVRDGTIKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSPK 1366
```

Qy 422 VCVNVTINPTQFNVPQNL 439
Db 1367 ECVNVTINPTQFNVPKNL 1384

Search completed: May 18, 2004, 11:42:44
Job time : 46.2755 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:37:00 ; Search time 11.4621 Seconds

(without alignments)
3684.135 Million cell updates/sec

Title: US-08-570-311-22

Perfect score: 2388

Sequence: 1 GTPNPNPNGTTLSSSF.....PKVCNVNTINPTQFNPVQNL 439

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 2388 | 100.0 | 2628 | 2 T28651 | hemagglutinin A - |
| 2 | 2058 | 86.2 | 1704 | 2 A55426 | gingipain R (EC 3. |
| 3 | 2037 | 85.3 | 1526 | 2 S49763 | gingipain R (EC 3. |
| 4 | 2025.5 | 84.8 | 1732 | 2 T30836 | lysine-specific cy |
| 5 | 251.5 | 10.5 | 991 | 2 I40229 | conserved hypotet |
| 6 | 136.5 | 5.7 | 1052 | 2 AF2959 | conserved hypotet |
| 7 | 136.5 | 5.7 | 1341 | 2 H98323 | autolysin [impor |
| 8 | 129.5 | 5.4 | 1248 | 2 C89874 | hypothetical prote |
| 9 | 129 | 5.4 | 691 | 2 B75622 | hypothetical prote |
| 10 | 124 | 5.2 | 5188 | 2 B85547 | probable RTX famil |
| 11 | 124 | 5.2 | 5291 | 2 P90696 | hypothetical prote |
| 12 | 123.5 | 5.2 | 1684 | 2 S10789 | amylase A-180 - al |
| 13 | 123 | 5.2 | 596 | 2 A55976 | cellulose 1,4-beta |
| 14 | 122.5 | 5.1 | 1090 | 2 S59077 | cellulose 1,4-beta |
| 15 | 121 | 5.1 | 1649 | 2 C86822 | hypothetical prote |
| 16 | 120.5 | 5.0 | 1345 | 2 H90975 | hypothetical prote |
| 17 | 119.5 | 5.0 | 2817 | 2 B97033 | uncharacterized pr |
| 18 | 119 | 5.0 | 715 | 2 JC4908 | alkaline serine pr |
| 19 | 118.5 | 5.0 | 1904 | 2 T13256 | tail-host specific |
| 20 | 118 | 4.9 | 1441 | 2 A86685 | prophage pil prote |
| 21 | 118 | 4.9 | 1939 | 2 D97316 | probable S-layer p |
| 22 | 118 | 4.9 | 4199 | 2 S76412 | hypothetical prote |
| 23 | 117.5 | 4.9 | 607 | 2 C69503 | conserved hypotet |
| 24 | 117 | 4.9 | 1122 | 2 T18346 | MGC1 protein precu |
| 25 | 116.5 | 4.9 | 1274 | 2 T10729 | transferrin-like p |
| 26 | 116.5 | 4.9 | 2660 | 2 E85822 | probable invasin Z |
| 27 | 116 | 4.9 | 4936 | 2 A82515 | hypothetical prote |
| 28 | 115.5 | 4.8 | 702 | 2 S48753 | major surface prot |
| 29 | 114.5 | 4.8 | 2468 | 2 A83412 | hypothetical prote |

30 114 4.8 1461 2 E90696
31 114 4.8 1461 2 A85547
32 112 4.7 872 2 S49541
33 112 4.7 1377 2 I54632
34 111.5 4.7 702 2 S48754
35 111.5 4.7 882 2 H82754
36 111 4.6 465 2 A47023
37 111 4.6 3624 2 AD0835
38 110.5 4.6 713 2 B75489
39 110.5 4.6 1268 2 AB0204
40 110.5 4.6 2044 2 AB1180
41 110.5 4.6 2554 2 AB3528
42 110 4.6 987 2 A64474
43 109.5 4.6 891 2 T36423
44 109.5 4.6 938 2 AF1772
45 109.5 4.6 1034 2 T30551

ALIGNMENTS

RESULT 1

T28651

hemagglutinin A - Porphyromonas gingivalis

C:Species: Porphyromonas gingivalis

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Aug-2001

C:Accession: T28651

R:Han, N.; Whitlock, J.; Progulake-Fox, A.

Infect. Immun. 64, 4000-4007, 1996

A:Title: The hemagglutinin gene A (hagA) of Porphyromonas gingivalis 381 contains four

A:Reference number: Z20494; MUID:97047672; PMID:8926061

A:Accession: T28651

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2628 <HAN>

A:Cross-references: EMBL:U41807; NID:gl552410; PID:gl469916; PIDN:AAB17128.1

C:Genetics:

A:Gene: hagA

| | | | | |
|-----------------------|-----------------|---|---------------|-------------------|
| Query Match | 100.0% | Score 2388; | DB 2: | Length 2628; |
| Best Local Similarity | 100.0% | Pred. No. 6.2e-159; | Mismatches 0; | Indels 0; Gaps 0; |
| Matches 439; | Conservative 0; | | | |
| QY | 1 | GTPNPNPNGTTLSSSFENGIPASWKTTDADGDNNTTTPPPGTSFAGHNSAICV | 60 | |
| DB | 1862 | GTPNPNPNGTTLSSSFENGIPASWKTTDADGDNNTTTPPPGTSFAGHNSAICV | 1921 | |
| QY | 61 | SSASVINFEQPNPNYLTPELSIPGGGTLTFWVCAODANYASEHYAVYASSTGNDASN | 120 | |
| DB | 1922 | SSASVINFEQPNPNYLTPELSIPGGGTLTFWVCAODANYASEHYAVYASSTGNDASN | 1981 | |
| QY | 121 | FANALLEVLTAKTVTVAPEAIRGRVQGTWVQKTVQLPAGTKYVAFRHFQCTDFFWNL | 180 | |
| DB | 1982 | FANALLEVLTAKTVTVAPEAIRGRVQGTWVQKTVQLPAGTKYVAFRHFQCTDFFWNL | 2041 | |
| QY | 181 | DEVEIKANGKRADETFETPESSTHGEAPAEWTTDADGGQGWCLSSGQDLWLTAGGTN | 240 | |
| DB | 2042 | DEVEIKANGKRADETFETPESSTHGEAPAEWTTDADGGQGWCLSSGQDLWLTAGGTN | 2101 | |
| QY | 241 | VVASFSWNGMALPNPNYLSKDVTCATKVKYVAVNDGFGPDGHYAVMLSKGTNAGDFTV | 300 | |
| DB | 2102 | VVASFSWNGMALPNPNYLSKDVTCATKVKYVAVNDGFGPDGHYAVMLSKGTNAGDFTV | 2161 | |
| QY | 301 | VFEETPNCKGKARFGLSTEADGAKPOSVMIERVLDLPAGTKYVAFRHYNCSDLNLYILL | 360 | |
| DB | 2162 | VFEETPNCKGKARFGLSTEADGAKPOSVMIERVLDLPAGTKYVAFRHYNCSDLNLYILL | 2221 | |
| QY | 361 | DDIQTMGSGSTPTDITYTVYVRDGTGFKIEGLTETTFEEDGVATGNHVCVVKYPAGVSP | 420 | |
| DB | 2222 | DDIQTMGSGSTPTDITYTVYVRDGTGFKIEGLTETTFEEDGVATGNHVCVVKYPAGVSP | 2281 | |
| QY | 421 | KVCNVNTINPTQFNPVQNL | 439 | |

Db 2282 KCVNVTINPTQNPVQNL 2300

RESULT 2
A55426
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis
N;Alternate names: 50K high molecular mass arginine-specific cysteine proteinase; HGP; R
C;Species: Porphyromonas gingivalis
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 08-Oct-1999
C;Accession: A55426; D53113
R;Pavloff, N.; Potempa, J.; Pike, R.N.; Prochazka, V.; Kiefer, M.C.; Travis, J.; Barr, R.
J. Biol. Chem. 270, 1007-1010, 1995
A;Title: Molecular cloning and structural characterization of the Arg-gingipain proteinase
A;Reference number: A55426; MUID:95138080; PMID:7846351
A;Accession: A55426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1704 <PAV>
A;Cross-references: GB:U15282; NID:g557067; PIDN:AAA69539.1; PID:g557068
B;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from Porphyromonas gingivalis. Isolat
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: D53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 228-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141694)
C;Keywords: cysteine proteinase; hydrolase

Query Match 86.2%; Score 2058; DB 2; Length 1704;
Best Local Similarity 88.8%; Pred. No. 4.9e-136; Mismatches 12; Indels 6; Gaps 4;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

Qy 3 PNPENPNPGTTLTSESFENGIPASWKTIADGNGNNWTTTPPGGTSPAGHNSAICVSS 62
Db 952 PNPENPNPGTTLTSESFENGIPASWKTIADGNGNGKPNAGP---IAGYNSGCVYS 1008
Qy 63 ASY-INFGPQNPNDYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1009 ESFGLGGIGVLTDPNLYITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASNF 1068
Qy 122 ANALLEEVLTAQVVTAPAIRGRVQGTWKVQKTVQVLAHFGCTDFFWNL 181
Db 1069 TNALLEETITAGK-VRSPEAIRG-RIQSTWRQKTVDLPAQTKYVAFRHFQSTDMFYIDL 1126
Qy 182 EVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDLTAHGGTNN 241
Db 1127 EVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDLTAHGGTNN 1186
Qy 242 VASPSWNGMALPNPNYLISKDVATGATKVKYKYAVNDGPGDHYAVMISKTGTNAGDFTVV 301
Db 1187 VASPSWNGMALPNPNYLISKDVATGATKVKYKYAVNDGPGDHYAVMISKTGTNAGDFTVV 1246
Qy 302 FEETPNGLKNGARGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNSDLNYILL 361
Db 1247 FEETPNGLKNGARGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNSDLNYILL 1306
Qy 362 DIQTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 421
Db 1307 DIQTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 1366
Qy 422 VCVNVTINPTQNPVQNL 439
Db 1367 ECVNVTINPTQNPVQNL 1384

RESULT 3
S49763
gingipain R (EC 3.4.22.37) precursor - Porphyromonas gingivalis (fragment)
C;Species: Porphyromonas gingivalis
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 31-Mar-1997

C;Accession: S49763
R;Aduse-Opoku, J.; Muir, J.; Slaney, J.M.; Rangarajan, M.; Curtis, M.A.
submitted to the EMBL Data Library, November 1994
A;Description: Cloning, sequence analysis and expression in Escherichia coli of prpA o:
A;Reference number: S49763
A;Accession: S49763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1526 <ADU>
A;Cross-references: EMBL:X82680
C;Genetics:
A;Gene: prpR1
C;Keywords: cysteine proteinase; hydrolase

Query Match 85.3%; Score 2037; DB 2; Length 1526;
Best Local Similarity 87.9%; Pred. No. 1.2e-134;
Matches 385; Conservative 13; Mismatches 34; Indels 6; Gaps 4;

Qy 3 PNPENPNPGTTLTSESFENGIPASWKTIADGNGNNWTTTPPGGTSPAGHNSAICVSS 62
Db 954 PNPENPNPGTTLTSESFENGIPASWKTIADGNGHGWKPNAGP---IAGYNSGCVYS 1010
Qy 63 ASY-INFGPQNPNDYLTPELSLPGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1011 ESFGLGGIGVLTDPNLYITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASNF 1070
Qy 122 ANALLEEVLTAQVVTAPAIRGRVQGTWKVQKTVQVLAHFGCTDFFWNL 181
Db 1071 TNALLEETITAGK-VRSPEAIRG-RIQSTWRQKTVDLPAQTKYVAFRHFQSTDMFYIDL 1128
Qy 182 EVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDLTAHGGTNN 241
Db 1129 EVEIKANGKRAADFTETFSSTHGEAPAEWTTIDADGGQGWLCSSGQLDLTAHGGTNN 1188
Qy 242 VASPSWNGMALPNPNYLISKDVATGATKVKYKYAVNDGPGDHYAVMISKTGTNAGDFTVV 301
Db 1189 VASPSWNGMALPNPNYLISKDVATGATKVKYKYAVNDGPGDHYAVMISKTGTNAGDFTVV 1248
Qy 302 FEETPNGLKNGARGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNSDLNYILL 361
Db 1249 FEETPNGLKNGARGLSTEADGAKPQSVWIERTVDLPAQTKYVAFRHYNSDLNYILL 1308
Qy 362 DIQTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 421
Db 1309 DIQTMGSSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHCEYCVKVTAGVSPK 1368
Qy 422 VCVNVTINPTQNPVQNL 439
Db 1369 KCVNVTINPTQNPVQNL 1386

RESULT 4
T30836
lysine-specific cysteine proteinase porphyrain (EC 3.4.22.-) - Porphyromonas gingivalis
N;Alternate names: lysine-specific cysteine proteinase 1, 60K
C;Species: Porphyromonas gingivalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30836; T30837; T30526; A53113
R;Barkocy-gallagher, G.A.; Han, N.; Patti, J.M.; Whitlock, J.; Progulski-Fox, A.; Lantz,
J. Bacteriol. 178, 2734-2741, 1996
A;Title: Analysis of the prpP gene encoding porphyrain, a cysteine proteinase of Porphy:
A;Reference number: 220895; MUID:96213011; PMID:8631659
A;Accession: T30836
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1732 <BAR>
A;Cross-references: EMBL:U42210; NID:g1314325; PID:g1314326; PIDN:AA06565.1
R;Slakeski, N.; Cleal, S.M.; Reynolds, E.C.
submitted to the EMBL Data Library, October 1996
A;Reference number: 220896
A;Accession: T30837
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA

A;Residues: 1-795, 'I', 797-1389, 'N', 1391-1478, 'Y', 1480-1732 <SLA>
A;Cross-references: EMBL:U75366; NID:G2182811; PID:G2182812; PIDN:AA60809.1
R;Lewis, J.P.; Macrina, F.L.
Infect. Immun. 66, 3035-3042, 1998
A;Title: IS195, an insertion sequence-like element associated with protease genes in *Porphyromonas gingivalis*. Isolated from strain 140229
A;Reference number: 220844; MUID:98298016; PMID:9632563
A;Accession: T30526
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1350, 'N', 1352-1363, 'Y', 1365-1447, 'H', 1449-1732 <LEW>
A;Cross-references: EMBL:AF017059; NID:G2738802; PID:G2738803; PIDN:AA26523.1
R;Pike, R.; McGraw, W.; Potempa, J.; Travis, J.
J. Biol. Chem. 269, 406-411, 1994
A;Title: Lysine- and arginine-specific proteinases from *Porphyromonas gingivalis*. Isolated from strain 140229
A;Reference number: A53113; MUID:94103245; PMID:8276827
A;Accession: A53113
A;Status: preliminary
A;Molecule type: protein
A;Residues: 229-249 <PIK>
A;Experimental source: H66
A;Note: sequence extracted from NCBI backbone (NCBIP:141690)
C;Genetics:
A;Gene: prtP; prtK
C;Keywords: cysteine proteinase; hydrolase

Query Match 84.8%; Score 2025.5; DB 2; Length 1732;
Best Local Similarity 87.1%; Pred. No. 9.4e-134;
Matches 384; Conservative 16; Mismatches 34; Indels 7; Gaps 5;
QY 1 GTNNPNPNPGT-TTILSESPENGIPASWKITDADGGQNNWTTTPPGCTSPAGHNSAIC 59
DB 969 GTNNPNPNPNPGT-TTILSESPENGIPASWKITDADGGHGWKPGNAPG---IAGYNSGC 1025
QY 60 VSSASY-INFEQPNPDNYLTPELSFGCGTGLFWCAQDANVASEHYAVASSTGND 118
DB 1026 VYSEFGLGGVLTPTDNYLTPLDENGKGLFWCAQDANVASEHYAVASSTGND 1085
QY 119 SNFANALLEEVLTAKTAVTAPEAIRTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFWI 178
DB 1086 SNFTNALLEETITAKG-VRSPEAIRG-RIOGTWRQKTVDLPAKTKYVAFRHFGQTDWFI 1143
QY 179 NLDEVEIKANGKRAADFETFPSSHTGEPAPAWTTIDADGGQWCLSSGGDLWLTAGG 238
DB 1144 DLDEVEIKANGKRAADFETFPSSHTGEPAPAWTTIDADGGQWCLSSGGDLWLTAGG 1203
QY 239 TNVVSFSSWNGMALNPONYLISKDVTGATKYKYAVYVNDGPPGDHYAVMISKTGTNAGDF 298
DB 1204 SNVVSFSSWNGMALNPONYLISKDVTGATKYKYAVYVNDGPPGDHYAVMISKTGTNAGDF 1263
QY 299 TVVFEETPNGKNGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHNGSDLYNI 358
DB 1264 TVVFEETPNGKNGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHNGSDLYNI 1323
QY 359 LLDDIQETMGSPPTDYTVVYRDGKIKKGLTETFEEDGVATGNHEYCVKVTAGV 418
DB 1324 LLDDIQETMGSPPTDYTVVYRDGKIKKGLTETFEEDGVATGNHEYCVKVTAGV 1383
QY 419 SPKVCVNVNTIPTQFNPQNL 439
DB 1384 SPKVCVNVNTIPTQFNPQNL 1404

RESULT 5
I40229
arginyl endopeptidase - *Porphyromonas gingivalis*
C;Species: *Porphyromonas gingivalis*
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40229
R;Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A;Title: Structural characterization of arginogipain, a novel arginine-specific cysteine
A;Reference number: I40229; MUID:95168884; PMID:7864651
A;Accession: I40229

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-991 <RES>
A;Cross-references: GB:D26470; NID:G927644; PIDN:BA005484.1; PID:G927645
Query Match 10.5%; Score 251.5; DB 2; Length 991;
Best Local Similarity 25.9%; Pred. No. 1.1e-09;
Matches 120; Conservative 50; Mismatches 171; Indels 123; Gaps 21;
QY 8 NNPNPGTTTILSESPENGIPASWKITDADGG-NNWT-----TTTPPP 46
DB 529 HPNNKRTFGVTNMGFMFAMVEKKYKGMKMLDWTVTGDPSSLVRLVPTKMGQVITAPQ 588
QY 47 GGTSTFAG-----HNSAICVSSASYINFEQPNPDNYLTPELS-LPGGGTTLTFWVCAQD 99
DB 589 INLTASVNVSCDYNGAIAIATISANGKF-GSAVVENGATINLTGLTNESTLITLV----- 643
QY 100 ANYASEHYAVASSTG--NDASNPANALLEEVLTAKTAVTAPEAIRTRVQGTWYQKTVQ 157
DB 644 VGYNKETVIKINTNGEPNPYPVSN-----LTATT-----QQKVTLKWDAPSTK 689
QY 158 LPAGTKYVAFRHFGCTDFWNL-DEVEIKANGKRAADFETFPSSHTGEPAPAWTTIDAD 216
DB 690 TNATTN-TARSVDGIRELVLLSVSDAPELLRSQGAIEVLEAHDVWNDG--GYQLLDAD 746
QY 217 GDGGQWCLSSGGDLWLTAGGTNVVASFSWN-----GMALNPONYLISKDVTGATKVKY 271
DB 747 HDQXGVIPSDTHTLWPCNSVPANLPAPFEYTVENADPSCSPNMM--DGTSVNI-- 802
QY 272 YVAVNDGPPGDHYAVMISKTGTNAGDTVVFEETPNGKNGARFGLSTEADGAKPQS-- 329
DB 803 -----PAGTY-----DFAI-----AAPQANA 818
QY 330 -VMT-----ERTVDLPAGTKYVAFRHNGYCVKVTAGVSPKVCVN 425
DB 819 KINTAGOPTYKEDDYVFEAKKY----HFLMKKSGSDGTELTISEGG-----SDYITTVY 871
QY 382 RDGFKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSPKVCVN 425
DB 872 RDGFKIKEGLTETTYRDAGSAQSHEVCVEKYAAGVSPKVCVD 915
RESULT 6
AF2959
conserved hypothetical protein Atu3276 [imported] - *Agrobacterium tumefaciens* (strain C
C;Species: *Agrobacterium tumefaciens*
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2959
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1052 <KUR>
A;Cross-references: GB:AE008689; PIDN:AA44092.1; PID:gl7741659; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3276
A;Map position: linear chromosome
Query Match 5.7%; Score 136.5; DB 2; Length 1052;
Best Local Similarity 21.1%; Pred. No. 0.14;
Matches 119; Conservative 61; Mismatches 148; Indels 235; Gaps 33;
QY 13 TTILSESPENGIPASWKITDADGGQNNWTT-----TTPPGTSPAGHNSAICVSSAS 64
DB 52 TVTSGEAIIGRHSAEATV--TGDGSKMTGTDQLQVGGTSDPGG--LAGNGT----- 99

QY 65 YINFEQPNPDNYLVTPELSLPGGSTLTFWVCAQDANYASEHYAVVASSTGN----- 116
Db 100 -----LNVTAGGVS-----DSTVA--HLGVVAGATGSAIVDGKGS 132
QY 117 -----DASNF-----ANALLEEVLTAKTAVTVAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 166
Db 133 VWTVDNRSLVGVSGAGSL--AVTGGGLVDAANIIGTNTG--NGSVRVSGADSTVK 186
QY 167 FRHFGCTDFWNL-----DEVEIKANG-----KRADFTETTESHGEAPAEWTIIDADGD 218
Db 187 SR-----SD-----LNVGLYNGSMTVEAGGAVKSRDGYVATYGGST-----SAVTVTGD 231
QY 219 GQWLCLSSGOLDWLTAGGT--NVVASFWSMGNALNPDNYLI-----SKDVTGA-TX 268
Db 232 GSWAMTGTFFGV--ASGATGNVTVS--NGGAIRATGVTGLDLAGASGTMITGAGSK 286
QY 269 VKYYY-----AVNDGFPG-----DHY-----AVMISKTGT-- 293
Db 287 VTAYVDNGTVNSGSDVDFGSGSLSVVNGSLDAYNLVYVGNALGSSGAVLVSGVSHVS 346
QY 294 -----NAGD-----FTVVPETPNKGGARFGLSTEADGAKPQSV 330
Db 347 VDGLMVVGNAGNSVEITGCGASLAAPTILIAATEAGSTGLSIGAGSGQTARSAGA----- 401
QY 331 WIE-RTVDLPAGTKYVAFRHYNCSDLYILLDDIQ-----FTMGSGPTPTDITY 378
Db 402 -VEARAIAFCAGNGSIVFNH--SETGYTILSADISAGRVVAEAGVTTLSGNNS----- 451
QY 379 TVYRDGTKIKEGLTETT-----FEEDG-----VA 402
Db 452 --YSGGTTISAGMLKGTAKSFGSGGIYVNAELVVDGGTILSNAISGTSPEKTGDGNLL 509
QY 403 TGNHEVCVEKYTAGVSPKVCVN 425
Db 510 TGNSTYSGATAVSAG---KLSVN 529

RESULT 7
H98323
hypothetical protein AGR_L_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: H98323
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H98323
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1341
A;Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:G15160106; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3085
A;Map position: linear chromosome

Query Match 5.7%; Score 136.5; DB 2; Length 1341;
Best Local Similarity 21.1%; Pred. No. 0.19; Indels 235; Gaps 33;
Matches 119; Conservative 61; Mismatches 148; Indels 235; Gaps 33;
QY 13 TTLSSEFENGIPASWKITDADGNNWTT-----TPPGGTSPAGHNSAICVSSAS 64
Db 341 TVTSGEALIGRHSASEATV--TGDGSKWTTGDLQVGGDTSDRGG--LAGNGT----- 388
QY 65 YINFEQPNPDNYLVTPELSLPGGSTLTFWVCAQDANYASEHYAVVASSTGN----- 116
Db 389 -----LNVTAGGVS-----DSTVA--HLGVVAGATGSAIVDGKGS 421
QY 117 -----DASNF-----ANALLEEVLTAKTAVTVAPEAIRGTRVQGTWYQKTVQLPAGTKYVA 166
Db 422 VWTVDNRSLVGVSGAGSL--AVTGGGLVDAANIIGTNTG--NGSVRVSGADSTVK 475

QY 167 FRHFGCTDFWNL-----DEVEIKANG-----KRADFTETTESHGEAPAEWTIIDADGD 218
Db 476 SR-----SD-----LNVGLYNGSMTVEAGGAVKSRDGYVATYGGST-----SAVTVTGD 520
QY 219 GQWLCLSSGOLDWLTAGGT--NVVASFWSMGNALNPDNYLI-----SKDVTGA-TX 268
Db 521 GSWAMTGTFFGV--ASGATGNVTVS--NGGAIRATGVTGLDLAGASGTMITGAGSK 575
QY 269 VKYYY-----AVNDGFPG-----DHY-----AVMISKTGT-- 293
Db 576 VTAYVDNGTVNSGSDVDFGSGSLSVVNGSLDAYNLVYVGNALGSSGAVLVSGVSHVS 635
QY 294 -----NAGD-----FTVVPETPNKGGARFGLSTEADGAKPQSV 330
Db 636 VDGLMVVGNAGNSVEITGCGASLAAPTILIAATEAGSTGLSIGAGSGQTARSAGA----- 690
QY 331 WIE-RTVDLPAGTKYVAFRHYNCSDLYILLDDIQ-----FTMGSGPTPTDITY 378
Db 691 -VEARAIAFCAGNGSIVFNH--SETGYTILSADISAGRVVAEAGVTTLSGNNS----- 740
QY 379 TVYRDGTKIKEGLTETT-----FEEDG-----VA 402
Db 741 --YSGGTTISAGMLKGTAKSFGSGGIYVNAELVVDGGTILSNAISGTSPEKTGDGNLL 798
QY 403 TGNHEVCVEKYTAGVSPKVCVN 425
Db 799 TGNSTYSGATAVSAG---KLSVN 818

RESULT 8

C89874
autolysin [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89874
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89874
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1248 <KUR>
A;Cross-references: GB:BA000018; PID:G13700854; PIDN:BA842150.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: atl

Query Match 5.4%; Score 129.5; DB 2; Length 1248;
Best Local Similarity 21.9%; Pred. No. 0.53;
Matches 103; Conservative 47; Mismatches 167; Indels 153; Gaps 25;
QY 3 PNEPNPNPCTTT---LSESEFENGIPASWKITDADGDC---NNWTTTPPP----- 46
Db 586 PTPTPIPKPSTPTNNKLTVSSLVG---AQINAKNGLFTTVYDKTKGPKTKVQKTF 641
QY 47 -----GCTSP---AGHNSAICV-----SSASYINFEQPN-PDNYLVTPELSL----- 85
Db 642 VTKEASLGNKFKYLVDYNSPTLIGWVKQGDVYNNAKSPVNMQTYTKPKGLYSVEV 701
QY 86 -----PGGTLTFWVCAQDANYASEHYAVVASSTGNDASNFAN-----ALLE 127
Db 702 GYKQAGAVSGTGNQTFKATKQQQIDKS---IYLFGTVNGKSGWVKAYLAVPAAPKK 757
QY 128 EVLTAKT-----VVTAPPAIR-----GTRVQGTWYQKTVQLPAGTKYVAFRHF 171
Db 758 AVAOPKTAVKYTVTKPQTTQVTSKIAOVKPNNTGIRASVIEKTAK--NGAKY-ADRTTY 814
QY 172 CT-----DFFWINLDEVEIKANGKRADFTE--TFESSTHGAPAE 209

Db 815 VTKERAHGNETVLLNNTSHNIPLGWENVKDLNVQNLGHEVKTOKYTVNKNNGLSMPV 874
QY 210 WTTIDADGGQWMLCLSSQDLWLAHGGTNVVASFSWN-----GMLNPNPNYLISKDV 263
Db 875 WGT-----KNQVILTGNNI-----AQGTFNATKQVSGKVLYLGTINNRGWNNAKDL 923
QY 264 TGATKVK-----YYAVNDGPPGDHYAVMISKTGT-----NAGDFTVVEETPN 307
Db 924 TAPTAKPTTSAADYNITYIVKNG-NGYIYVTPNSDTAKYSLKAFNEQFFAVVKEQVIN 982
QY 308 G-----INKGARGFLSTEADGAK-----PQSVWIERTVDLPAGTKY 344
Db 983 GGTWYGLKSLGKLAWIKST--DLAKELIKYQNTGMLNQVAIQGLQY 1030
RESULT 9
B75622
Hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75622
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: B75622
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-691 <WHI>
A:Cross-references: GB:AE001826; NID:G6460827; PIDN:AAF12628.1; PID:G6460924; TIGR:DRB00
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRB0037
A:Map position: megaplasmid
A:Genome: plasmid
A>Note: plasmid MPI
Query Match 5.4%; Score 129; DB 2; Length 691;
Best Local Similarity 21.3%; Pred. No. 0.27;
Matches 113; Conservative 52; Mismatches 192; Indels 174; Gaps 24;
QY 4 NNPNPNPQTTLSESFENGIPASWKTIADGDGNNWTTT-----PPGGTSF 51
Db 230 SPNPLPPTGTTT-----PQTGTPVSGNGPAESTGPATGTGIGTSGTQPPVYVDA 280
QY 52 AGHNSAICVSSASYINFEGP-----QNPD-----NVLVTPELS-LPGGGTLTFWVCAQDA 100
Db 281 SGQVAYPADAEVTPNSVTMTIITNPNPAATVELVDLSGLPAGVTVTF-----TDA 336
QY 101 N-----YASBHYAVYASSTGNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQ 153
Db 337 NGNPLPDTGNGRPEVTAGNGTA-----TYRVVTVYPTDESAAVAG-----380
QY 154 KTVQLPAGTK-----YVAFHFCTDFFINLDEVEIKANGKRADT-----195
Db 381 -PIRIPVGVGNGRDIGVDAVTYVNVLLSNLKFNGTNGTALGVSDVPVTVTTPQSPVTTAV 439
QY 196 ----ETFESSSTH-----GEAPAEWTITDADGGQGWLCSSGOLDWLTAAHGT 239
Db 440 VFPMDLNDGAYDGNVALSGSTPIGPVKYATNPDTDGG-----VLSPAEALAPAE-----492
QY 240 NVVASFSWNGMALPNPNYILSKDVTGATKVKYVAVNDGPPGDHYAVMISKTGTAGDFT 299
Db 493 --IAS-----TGAVPVKTEKT-----YAVVTIPAGQAGDYM 523
QY 300 VVEETPNKGGARFGLSTEADGAKPO-SVWIERTVDLPAGT-----342
Db 524 VT--QTATGSLSGTTKTSFNTKVTVTPSNGSLIIAKRVTTTGTTPSLNATANPGDAVS 581
QY 343 KYVAFRHYNCSDLNILLDDIQTMTG-----GSPTPTDYTYTV-----YRD-- 383

Db 582 TVATNNYNTSLYGLVLRDPSSNNLGSFSSNVFGFIKPSLRATVSGVSGATVLYRTSN 641
QY 384 -GTKIKEGLTE--TTFEEDGVATGNHNEVCVEKYTAGVSPKVCVNVNTIPT 431
Db 642 LNTWAAQPTVDANTTWVEGVDTNNNS---QIDSGDVFPFNAVITLTLOGT 689
RESULT 10
B85547
Probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:G12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:ZC
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615
Query Match 5.2%; Score 124; DB 2; Length 5188;
Best Local Similarity 23.9%; Pred. No. 8.4;
Matches 138; Conservative 63; Mismatches 189; Indels 188; Gaps 35;
QY 10 NPGTTL--SESFENGIPAS-----WKTIDA---DGDGNNWTTT-----PPEG 47
Db 2180 NTFTTLDASGNNVSGVPASVVSALANGTVTINASVTDAGNSGSATHQVTVNTGLPTIT 2239
QY 48 GTSPAGN-----SAICVSSASYINFEGPQ-----NPNLYLTPELSLPGGTLTFW 94
Db 2240 FNAISGDNILNADEKQPLTISGGSTGLATGAQVTVTLNGHNSATTDAS--GNWTLTVP 2297
QY 95 VCAQDANYASEHYAVYASST---GNDASNFANALLEEVLTAKTVV-----A 138
Db 2298 V-SDLAALQOANYTVSASATSAAGNTASSQANLLVDSGLPDVTINTVAGDDIINAAEAG 2356
QY 139 PEAIRG--TR-----VOGTWYQKTVQLPAGTKYVAFRHFGCTDFFW-INLDEVEI 185
Db 2357 DQTSISGVVTRAAAGDVTVTTLGNTYTAIVQ-----SNLSWSVSVPTADL 2401
QY 186 KANGKRADETFETFEESTHGEAPAEWT---TIDAD-----GD-----220
Db 2402 QALG-NGDLTITFASVINGANGTSGSTRDITIDANLFGURVDVTVAGDDIVNSTEHGQALVI 2460
QY 221 --GWLCLSGQLDMLTAHG--GTNVVASFSWN-GM-ALNPDNY---LISKDVTGATKVK 270
Db 2461 TCGSSCLNAGAVLTVINSVAYSATVQADGSHVSGVIPAANVSAWPAAGPLTVEVDGSSAN 2520
QY 271 YYAVNDGPPGDHYAVMIS-----KTGTNAGDFTVVEETPNINK-----311
Db 2521 NPVSVSHPTVDLTAVAISINTVASDDVINAEEKGTN-----LTLGSGTSGIESGTQTVV 2575
QY 312 --GGARFGLSTEADGAKPQSVWIERTVDLPAGTKVK-----AFRHYNC-----353
Db 2576 TFGGKTYTASVAANGSWNVNPAADLTLPEGAANVQASVSSAGNSASATHAYSVDASA 2635
QY 354 ---DLNLYILLDDI-----QFTMGSGPT-PTDYTYVYRDGTIKKEGLTET-TFEEDG 400
Db 2636 PLTINTASDILNAAEAGSLTISGTSTAGTGTQTVTLNGA-----TYTGTQADG 2689
QY 401 VATGNEHYCVEKYTA-GV--SPKVCVNVNTINPTQNP 435
Db 2690 -----SWSVSVPTSGALNALNASVTVSATVNDKAGNP 2721
RESULT 11

F90696
hypothetical protein EC0542 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: CB:BA000007; PIDN:BA033965.1; PID:G13360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: EC0542

Query Match 5.2%; Score 124; DB 2; Length 5291;
Best Local Similarity 23.9%; Pred. No. 8.6;
Matches 138; Conservative 63; Mismatches 189; Indels 188; Gaps 35;

QY 10 NPGTTTL-SESFNGIPAS-----WKTTDA---DGGNNWTT-----PPG 47
DB 2180 NTFTVLDSGNWSGVGPASVVSALANGTVTINASVTDAGSGSATHQVTNTGLPTIT 2239
QY 48 GTSFAGHN-----SAICVSSASYINFEPQ-----NPDNLYTPELSIPGGGTLTFW 94
DB 2240 FNAISGDNILNADKGOPLTISGSGTGLATGAQVTVTLNGHNSATTDAS--GNWTLTVP 2297
QY 95 VCAQDANYASBYHYVYASST---GNDASNFANALLEEVLTAKTAVT-----A 138
DB 2298 V-SDLAALGQANYTVSASATSAAGNTASSQANLLVDSGLPVDVTINTVAGDDIINAEAGA 2356
QY 139 PEARG--TR-----VOGTWYQKTVLPAGTKYVAFRHFCTDFF--WINLDEVEIKANGRADFTE 196
DB 2357 DQTTISGVVTRAAAGDTVTTLGGNTYTATVQ-----SNLSWSVSVPATDL 2401
QY 186 KANGKRAADFTTFESSTHGEPAEW---TTDAD-----GD-----GO----- 220
DB 2402 QALG-NGDLTITASVTNANGNGGTRDITDANLPLGRVTVAGDDIIVNSIEHQALVI 2460
QY 221 --GWLCLSSGQDLWLTAHG---GTNVASFVN--GM-ALNPDNY---LISKDVTGATKVK 270
DB 2461 TGGSSGLNAGAVLTVTINSVAYSATVQADGWSVGIPAAVNSAWPAGPLTVEVDGQSSAN 2520
QY 271 YYYAVNDGPPGDHYAVMIS-----KGTNAGDFTVVPFEETPENGK----- 311
DB 2521 NPVSVSHPFTVLDLTAVALISINTVASDDVINAABKGTN-----LTLGSGTSGIESGQTVTV 2575
QY 312 --GGARFGLSTEADGAKPQSVIERTVLDLPAGTKV-----AFRHYNCS--- 353
DB 2576 TFGKTYTASVANGSWSVNPPAADLATLPEGAANVQASVSASGNSASATHAVSVDASA 2635
QY 354 ---DLNYLLDDI-----QFTMGSGPT-PTDYTVTVYRDGKIKKGLTET--TFEEDG 400
DB 2636 PTLTINTIASDILNAAEAGSPLATISGTSTAETGQTVTVTLNGA-----TYTGTVQADG 2689
QY 401 VATGNHCVKYEKTA-GV--SPKVCUNVTINPTQNP 435
DB 2690 -----SWSVSPTSALGALNASNTVTSATVNDKAGNP 2721

RESULT 12
S10789
amylase A-180 - alkaliphilic eubacterium 163-26
C:Species: alkaliphilic eubacterium 163-26
C:Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
C:Accession: S10789
R: Candussio, A.; Schmid, G.; Boeck, A.
Eur. J. Biochem. 191, 177-185, 1990
A:Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an

A:Reference number: S10789; MUID:903336627; PMID:1696201
A:Accession: S10789
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1684 <CAN>
A:Cross-references: EMBL:X53373; NID:948305; PIDN:CAA37453.1; PID:948306

Query Match 5.2%; Score 123.5; DB 2; Length 1684;
Best Local Similarity 20.7%; Pred. No. 2.1;
Matches 100; Conservative 52; Mismatches 183; Indels 149; Gaps 27;

QY 23 GIPASWKTIDADGDNWTTTPPPGTSFAGHNSAICV-SSASYINPEGQNDNVLVTP 81
DB 204 GLPRDWTNQA--QGQNWHT-----HNDIMNKONEAAAWANWWSGD----- 241
QY 82 ELSLPGGGTITFWVCAQD-ANYASEHYAVYASSTGNDASNFANALLEEVLTAKTV-VTAP 139
DB 242 -----WIRADETAGY-----DCCGSEQTMCIQGLPDINKTEVTTGVLDLP 280
QY 140 EAIRGT-RVQGTWYQKTVLPAGTKYVAFRHFCTDFF--WINLDEVEIKANGRADFTE 196
DB 281 PILRNKWDQASGYEDWF-VPAAEPYRQDLINIAPKDYLIKWITSWVEEFGIDGRVDTAK 339
QY 197 TFSSTHGEAPAB-----WTTIDADGQGMCLSSGQDLWLTA-----HG----- 237
DB 340 HVEIERMAELKNEAEVALQWRENNPDKPCANW-----DDNFMTAEVFGHGLKSEYFD 394
QY 238 -GTNVASFWSW-----NGMALPDNLYLSKDVGTGATKVKYVYAVNDGPP 280
DB 395 FGDSVINFEFQANFNENLGLFSRVANSINTDPDNMLSYVSSHDTKL---YSRDD--- 448
QY 281 GDHYAVMISKTGTN---AGDFTVVF--EETPNKNGGARFGLSTEADGAKPOSVM--IE 333
DB 449 -----LIQAGTALLLPQGVQVFGDETAFLPGDGG-----SDPEQGTSSNMWANIN 496
QY 334 RTVDLPAGTKYVAFR-----HYNCSDLNVI-----LDDDIQFTWGGSTPT 374
DB 497 QNV-LSHWQKLQGOFRNNHIAIGAGAHQKLSDSPTFARTYESDDIDVDEVVATGAQGTTA 555
QY 375 DYTTVYRDGKIKKGLT--ETTFEEDGATGNHCVKYEKTAAGSPKVCUNVTINPTQ 432
DB 556 VTVEGVFEDGTVRDVTGDETTVTK-GTAT-----FRTAGTQGIILIENTASPTV 604
QY 433 FNPV 436
DB 605 NLPI 608

RESULT 13
A55976
cellulose 1,4-beta-cellubiosidase (EC 3.2.1.91) - Thermomonospora fusca
C:Species: Thermomonospora fusca
C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
C:Accession: A55976
R: Zhang, S.; Lao, G.; Wilson, D.B.
Biochemistry 34, 3386-3395, 1995
A:Title: Characterization of a Thermomonospora fusca exocellulase.
A:Reference number: A55976; MUID:95186496; PMID:7880834
A:Accession: A55976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-596 <ZHA>
A:Cross-references: GB:U18978; NID:9664822; PIDN:AAA62211.1; PID:9664823
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:40-140/Domain: bacterial cellulose-binding domain homology <BCB>
F:41-139/Disulfide bonds: #status predicted

Query Match 5.2%; Score 123; DB 2; Length 596;
Best Local Similarity 23.2%; Pred. No. 0.58;
Matches 69; Conservative 35; Mismatches 112; Indels 82; Gaps 14;

QY 3 PNPFPNPFGTTLTSESFENGIPASWKT-----IDAGDGNWTTTP-----PFGGT 49

Db 450 GQALFTGNGAGTGAQA-----TNSTTAAQGILYANITNVIKSAFTTGTAG 497

QY 421 KVC-----VNVITNPTQF 433

Db 498 AVYGIYGGNGHDSLKISPSQW 518

Search completed: May 18, 2004, 11:47:52
Job time : 13.4621 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:35:14 ; Search time 8.0235 Seconds
(without alignments)
2848.981 Million cell updates/sec

Title: US-08-570-311-22

Perfect score: 2388

Sequence: 1 GTPNPNPNGPQTTLSESF.....PKVCNVNTINPTQNFVQNL 439

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 2388 | 100.0 | 2628 | 1 HGA2_PORGI | Q51845 porphyromon |
| 2 | 2335 | 97.8 | 2164 | 1 HGA2_PORGI | P59915 porphyromon |
| 3 | 1431 | 59.9 | 989 | 1 PRTH_PORGI | P46071 porphyromon |
| 4 | 251.5 | 10.5 | 931 | 1 CPGI_PORGI | P28784 porphyromon |
| 5 | 132.5 | 5.5 | 1256 | 1 ATL_STAUF | P52081 staphylococ |
| 6 | 122.5 | 5.1 | 1090 | 1 GUXB_CELFI | P50899 cellulomona |
| 7 | 120.5 | 5.0 | 2660 | 1 YEEJ_ECO57 | Q8x8v7 escherichia |
| 8 | 117.5 | 4.9 | 331 | 1 PME_ASPAC | Q12535 aspergillus |
| 9 | 117.5 | 4.9 | 607 | 1 YK28_ARCFU | O28251 archaeoglob |
| 10 | 117 | 4.9 | 1122 | 1 ADP1_MYCGA | Q49379 mycoplasma |
| 11 | 113 | 4.7 | 1122 | 1 ADP2_MYCGA | Q9rem8 mycoplasma |
| 12 | 112.5 | 4.7 | 1260 | 1 ALS1_CANAL | P46590 candida alb |
| 13 | 112 | 4.7 | 872 | 1 GUXA_CELFI | P50401 cellulomona |
| 14 | 111 | 4.6 | 465 | 1 SLAP_LACBR | O05044 lactobacill |
| 15 | 110 | 4.6 | 987 | 1 YD94_METUA | Q58789 methanococ |
| 16 | 109.5 | 4.6 | 1385 | 1 C5AA_BACUD | Q45760 bacillus th |
| 17 | 107.5 | 4.5 | 1034 | 1 ITAV_CHICK | P26008 gallus gall |
| 18 | 107.5 | 4.5 | 1200 | 1 HYAL_STRPU | O76536 strongyloce |
| 19 | 107.5 | 4.5 | 1953 | 1 BIGA_SALTU | P25927 salmonella |
| 20 | 107 | 4.5 | 435 | 1 AM3D_ORYSA | P27933 oryza sativ |
| 21 | 105.5 | 4.4 | 282 | 1 PRTA_ASPNG | P24655 aspergillus |
| 22 | 105.5 | 4.4 | 699 | 1 CH11_BACCI | P20533 bacillus ci |
| 23 | 105.5 | 4.4 | 1609 | 1 FIG2_YEAST | P25653 saccharomyc |
| 24 | 105 | 4.4 | 524 | 1 CH1D_BACCI | P27050 bacillus ci |
| 25 | 105 | 4.4 | 2812 | 1 ZAN_HUMAN | O9y493 homo sapien |
| 26 | 104.5 | 4.4 | 551 | 1 AMYB_THETU | P19584 thermoaer |
| 27 | 104.5 | 4.4 | 1034 | 1 BGAL_BACME | O52847 bacillus me |
| 28 | 104 | 4.4 | 721 | 1 OGP_MOUSE | Q62010 mus musculu |
| 29 | 104 | 4.4 | 1443 | 1 NEOI_CHICK | Q90610 gallus gall |
| 30 | 103 | 4.3 | 548 | 1 THER_BACST | P06874 bacillus st |
| 31 | 102.5 | 4.3 | 527 | 1 NPPE_BACBR | P43263 bacillus br |
| 32 | 102.5 | 4.3 | 1220 | 1 C5AC_BACTU | P56955 bacillus th |
| 33 | 102 | 4.3 | 1044 | 1 ITA8_CHICK | P26009 gallus gall |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 102 | 4.3 | 1045 | 1 GUNB_CELFI | P26225 cellulomona |
| 35 | 102 | 4.3 | 1289 | 1 CSAB_BACUD | Q45753 bacillus th |
| 36 | 101.5 | 4.3 | 837 | 1 XYNZ_CLOTM | P10478 clostridium |
| 37 | 101.5 | 4.3 | 2124 | 1 PGCA_RAT | P07897 rattus norv |
| 38 | 101.5 | 4.3 | 2358 | 1 YEEJ_ECOLI | P76347 escherichia |
| 39 | 101 | 4.2 | 439 | 1 SLAP_LACHE | P38059 lactobacill |
| 40 | 101 | 4.2 | 827 | 1 XANP_XANS2 | Q60106 xanthomonas |
| 41 | 100.5 | 4.2 | 1377 | 1 NSOI_RAT | P97603 rattus norv |
| 42 | 100.5 | 4.2 | 2249 | 1 OMPA_RICRI | P15921 rickettsia |
| 43 | 100 | 4.2 | 623 | 1 T3BE_BPSFV | Q9xip3 bacterioph |
| 44 | 100 | 4.2 | 762 | 1 E13B_TRIHA | P53626 trichoderma |
| 45 | 100 | 4.2 | 1157 | 1 C8AA_BACUK | Q45704 bacillus th |

ALIGNMENTS

```

RESULT 1
HGA2_PORGI
ID HGA2_PORGI STANDARD; PRT; 2628 AA.
AC Q51845;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin A precursor.
GN HGA.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=381;
RX MEDLINE=97047672; PubMed=8926061;
RA Han N., Whitlock J., Progulskie-Fox A.;
RT "The hemagglutinin gene A (hga) of Porphyromonas gingivalis 381
contains four large, contiguous, direct repeats.";
RL Infect. Immun. 64:4000-4007(1996).
CC -!- FUNCTION: Agglutinates erythrocytes.
CC -!- SIMILARITY: Belongs to peptidase family C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; U41807; AAB17128.1; -
PIR; T28651; T28651.
KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 2628 HEMAGGLUTININ A.
FT DOMAIN 25 539 PEPTIDASE C35-LIKE 1.
FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.
FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.
FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.
FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.
SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;
Query Match 100.0%; Score 2388; DB 1; Length 2628;
Best Local Similarity 100.0%; Pred. No. 1.6e-160;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTPNPNPNGPQTTLSESFENGIPASWKITDADGQNNWTTTPPGGTSGFAGNKAICV 60
Db 1862 GTPNPNPNGPQTTLSESFENGIPASWKITDADGQNNWTTTPPGGTSGFAGNKAICV 1921
Qy 61 SSASYINFEQPNPDNYLVTPELSFGGGLTFVWCAQDANYASEHYAVYASSTGNDASN 120
Db 1922 SSASYINFEQPNPDNYLVTPELSFGGGLTFVWCAQDANYASEHYAVYASSTGNDASN 1981

```

QY 121 FANALIEVLTAKTAVTAPEAIRTRVQGTWYQKTVOLPAGTKYVAPRHFGCTDFFWINL 180
 DB 1982 FANALIEVLTAKTAVTAPEAIRTRVQGTWYQKTVOLPAGTKYVAPRHFGCTDFFWINL 2041
 QY 181 DEVEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDMLTAHGGTN 240
 DB 2042 DEVEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDMLTAHGGTN 2101
 QY 241 VVASFSGWNGMALNDPNYLISKDVTGATKVKYIYVANDGPGPDHYAVMISKTGTNAGDFTV 300
 DB 2102 VVASFSGWNGMALNDPNYLISKDVTGATKVKYIYVANDGPGPDHYAVMISKTGTNAGDFTV 2161
 QY 301 VFEETPGINKGARFGLSTEDGAKPQSVIERTVDLPAGTKYVAPRHYNCSDLNILL 360
 DB 2162 VFEETPGINKGARFGLSTEDGAKPQSVIERTVDLPAGTKYVAPRHYNCSDLNILL 2221
 QY 361 DDIOFTMGSGPTPTDYTYTVYRDGTIKIKEGLTETFEEDGVATGNHEYCVVEKVTAGVSP 420
 DB 2222 DDIOFTMGSGPTPTDYTYTVYRDGTIKIKEGLTETFEEDGVATGNHEYCVVEKVTAGVSP 2281
 QY 421 KVCVNTINPTQFNPVQNL 439
 DB 2282 KVCVNTINPTQFNPVQNL 2300

RESULT 2

HGAL_PORGI STANDARD; PRT; 2164 AA.
 ID HGAL_PORGI
 AC P53915;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hemagglutinin A precursor.
 GN HAGA OR PG1837.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 CX NCBI_TaxID=837;
 RN [1]
 RC STRAIN=W83;
 RX MEDLINE=22829867; PubMed=12949112;
 RA Nelson K.E., Fleischmann R.D., Deboy R.T., Paulsen I.T., Fouts D.E.,
 RA Eisen J.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.,
 RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T., Tallon L., Gray J.,
 RA Granger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
 RA Dewhirst F.E., Fraser C.M.;
 RT "Complete genome sequence of the oral pathogenic bacterium
 RT Porphyromonas gingivalis strain W83."
 RL J. Bacteriol. 185:5591-5601(2003).
 CC -!- FUNCTION: Agglutinates erythrocytes (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family C25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AB017178; AAQ66831.1; ALT_INIT.
 DR TIGR; PG1837;
 KW Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2164 HEMAGGLUTININ A.
 FT DOMAIN 26 539 PEPTIDASE C25-LIKE 1.
 FT DOMAIN 540 991 PEPTIDASE C25-LIKE 2.
 FT DOMAIN 992 1443 PEPTIDASE C25-LIKE 3.
 SQ SEQUENCE 2164 AA; 233387 MW; 6DFAB22832586C63 CRC64;

Query Match

97.8%; Score 2335; DB 1; Length 2164;

Best Local Similarity 98.4%; Pred. No. 6.9e-157;
 Matches 430; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PNPENPNPGCTTLLSSFPENGIPASWKTIADADGGNNWTTTPPGGTSFAGHNSAICVSS 62
 DB 1400 PNPENPNPGCTTLLSSFPENGIPASWKTIADADGGNNWTTTPPGGTSFAGHNSAICVSS 1459
 QY 63 ASYINTEGPNPNPNYLVTPELSPPGGTTLTFWCAODANYASBHYAVYASSTGNDASNFA 122
 DB 1460 ASYINTEGPNPNPNYLVTPELSPPGGTTLTFWCAODANYASBHYAVYASSTGNDASNFA 1519
 QY 123 NALLEEVLTAKTAVTAPEAIRTRVQGTWYQKTVOLPAGTKYVAPRHFGCTDFFWINLDE 182
 DB 1520 NALLEEVLTAKTAVTAPEAIRTRVQGTWYQKTVOLPAGTKYVAPRHFGCTDFFWINLDD 1579
 QY 183 VEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDMLTAHGGTNV 242
 DB 1580 VEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDMLTAHGGTNV 1639
 QY 243 ASFSWNGMALNDPNYLISKDVTGATKVKYIYVANDGPGPDHYAVMISKTGTNAGDFTVWF 302
 DB 1640 ASFSWNGMALNDPNYLISKDVTGATKVKYIYVANDGPGPDHYAVMISKTGTNAGDFTVWF 1699
 QY 303 EETPGINKGARFGLSTEDGAKPQSVIERTVDLPAGTKYVAPRHYNCSDLNILLDD 362
 DB 1700 EETPGINKGARFGLSTEDGAKPQSVIERTVDLPAGTKYVAPRHYNCSDLNILLDD 1759
 QY 363 IOFTMGSGPTPTDYTYTVYRDGTIKIKEGLTETFEEDGVATGNHEYCVVEKVTAGVSPKV 422
 DB 1760 IOFTMGSGPTPTDYTYTVYRDGTIKIKEGLTETFEEDGVATGNHEYCVVEKVTAGVSPKE 1819
 QY 423 CVCVNTINPTQFNPVQNL 439
 DB 1820 CVCVNTINPTQFNPVQNL 1836

RESULT 3

PRTH_PORGI STANDARD; PRT; 989 AA.
 ID PRTH_PORGI
 AC P46071;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protease prth (EC 3.4.22.-).
 GN PRTH.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 CX NCBI_TaxID=837;
 RN [1]
 RC STRAIN=W83;
 RX MEDLINE=95012612; PubMed=7927685;
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RT "Cloning and characterization of a new protease gene (prth) from
 RT Porphyromonas gingivalis."
 RL Infect. Immun. 62:4279-4286(1994).
 RN [2]
 RP ERRATUM.
 RA Fletcher H.M., Schenkein H.A., Macrina F.L.;
 RL Infect. Immun. 62:5707-5707(1994).
 CC -!- FUNCTION: CLEAVES HUMAN COMPLEMENT COMPONENT C3. MAY ENABLE
 CC P.GINGIVALIS TO EVADE COMPLEMENT-MEDIATED KILLING DURING THE
 CC IMMUNE RESPONSE. PLAYS AN IMPORTANT ROLE IN SOFT TISSUE INFECTIONS
 CC AND IS A VIRULENCE FACTOR.
 CC -!- SUBCELLULAR LOCATION: In membrane vesicles
 CC -!- SIMILARITY: Belongs to peptidase family C25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL; L27483; AAA51298.1; -.
CC DR HSP; F23882; 1FMT.
CC DR MEROPS; C25.001; -.
CC DR InterPro; IPR002376; formyl transf.
CC DR Pfam; PF00551; formyl transf; 1.
CC KW Hydrolase; Thiol protease; Repeat; Virulence.
CC FT REPEAT 270 323
CC FT REPEAT 528 581
CC SQ SEQUENCE 989 AA; 110238 MW; FA85FE8A3AC8944C CRC64;

Query Match
Best Local Similarity 59.9%; Score 1431; DB 1; Length 989;
Matches 271; Conservative 8; Mismatches 21; Indels 8; Gaps 3;

QY 136 VTAPAIRTRVQGTWYQKTVQVLPAGTKYVAFRHFG-----CTDFWINLDEVEIKANGKR 191
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 34 VRSPEAIRG-RIGTWRQKTVLPAGTEICCFPSLPKAPICST---STLMRLRSKTNKR 89
QY 192 ADFTTFESSTGEAPAEWTTIDADGGGCLSSGQDLWLTAGGTNNVASFSGWNGMA 251
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 90 ADFTTFESSTGEAPAEWTTIDADGGGCLSSGQDLWLTAGGTNNVASFSGWNGMA 149
QY 252 LNPNLYISKVGTGATKVKYVAVNDGPFDDHYAVMISKGTNAGDFTVVFEETNGINK 311
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 150 LNPNLYISKVGTGATKVKYVAVNDGPFDDHYAVMISKGTNAGDFTVVFEETNGINK 209
QY 312 GGARGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHVNCSDNLYLLDDIQFTMGSSP 371
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 210 GGARGLSTEADGAKPQSVIERTVDLPAGTKYVAFRHVNCSDNLYLLDDIQFTMGSSP 269
QY 372 TPTDVTYTVRGTGKIKEGLTETTFEDGVAIGHEVCYEVKTAGVSKVCNVNTINPT 431
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 270 TPTDVTYTVRGTGKIKEGLTETTFEDGVAIGHEVCYEVKTAGVSKVCNVNTINPT 329
QY 432 QFNPNVQNL 439
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 330 QFNPNVQNL 337

RESULT 4
ID CPGL PORGI STANDARD; PRT; 991 AA.
AC P28784; Q45168;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Gingipain R1 precursor (EC 3.4.22.37) (Gingipain 1) (Arg-gingipain)
DE (RGP-1).
GN RGA OR RGP1.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 228-230 AND 517-541.
RC STRAIN=381;
RX MEDLINE=95168884; PubMed=7864651;
RA Okamoto K., Misumi Y., Kadowaki T., Yoneda M., Yamamoto K.,
RA Ikehara Y.;
RT "Structural characterization of arginipain, a novel
RT arginine-specific cysteine proteinase as a major periodontal
RT pathogenic factor from Porphyromonas gingivalis.";
RL Arch. Biochem. Biophys. 316:917-925 (1995).
RN [2]
RP SEQUENCE OF 228-270.
RC STRAIN=HG66;
RX MEDLINE=92406812; PubMed=1527017;
RA Chen Z., Potempa J., Polanowski A., Wikstrom M., Travis J.;
RT "Purification and characterization of a 50-kDa cysteine proteinase
RT (gingipain) from Porphyromonas gingivalis.";

```

```

Query Match
Best Local Similarity 10.5%; Score 251.5; DB 1; Length 991;
Matches 120; Conservative 50; Mismatches 171; Indels 123; Gaps 21;

QY 8 NPNPGTTLSEFENGIPASWKTIDADGG--NNMT-----TTPPP 46
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 529 HNNIKRIFGGVTMGMFAWVEKYKDGKLDWTVFGDPSLLVRLVPTKMQVTAPAQ 588
QY 47 GGTSPAG-----HNSAICVSSASVINFGPNPNLYLTPELS-LPGGGTLTFWVCAQD 99
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 589 INLTASVNVSCDYNGAATISANGKMF--GSAAVNGTATINLTGLTNESTLTAV--- 643
QY 100 ANYASEHYAVYASSTG--NDASNEANALLEEVLTAKTAVTAPEAIRGTRVQGTQKTVQ 157
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 644 VGNETVTKTINTNGEPNYPQVSN-----LTATT-----GQKVTLKNAPSTK 689
QY 158 LPAGTKYVAFRHFGCTDFWINL-DEVEIKANGKADFTTETPESSTHSGAPAEWTTIDAD 216
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 690 TNATTN-TARSVDGIRELVLLSVSDAPELRSSQAEIVLEAHDVWNGDS--GVQILLDAD 746
QY 217 GGGQWCLSSGQDLWLTAGGTNNVASFSGW-----GNALPNPNLYISKVGTGATKVKY 271
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 747 HDQYGVIPSDTHTLWPNCSVPEANLFAPEYTVPEADPSCSFTNMIM--DGTASVNI-- 802
QY 272 YVAVDGGPDGHYAVMISKGTNAGDFTVVFEETNGINKGARGCLSTEADGAKPQS-- 329
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 803 -----PAGTY-----DFAI-----AAPQANA 818
QY 330 -VMI-----ERTVDLPAGTKYVAFRHVNCSDNLYLLDDIQFTMGSSPTTDTYTYV 381
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 819 KIWIAGQGTKEDDYVFEAGKY----HPLMKMGSGDGTTLTISEGGG--SDYTYTYV 871

```

J. Biol. Chem. 267:18896-18901(1992).
 !- FUNCTION: Thiol protease which is believed to participate in intracellular degradation and turnover of proteins. Its proteolytic activity is a major factor in both periodontal tissue destruction and in bacterial host defense mechanisms. Activates complement C3 and C5.
 !- CATALYTIC ACTIVITY: Cleavage of proteins, including collagens and immunoglobulins, with a preference for Arg in P1, and hydrophobic residues in P2 and P3.
 !- ENZYME REGULATION: Requires cysteine for activation and Ca(2+) and/or Mg(2+) for stabilization. It is stimulated by glycine-containing dipeptides. It is resistant to inhibition by proteinase inhibitors in human plasma.
 !- SIMILARITY: Belongs to peptidase family C25.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

EMBL; D26470; BAA05484.1; -.
PIR; I40229; I40229.
DR HSP; P95493; 1CVR.
DR MEROPS; C25.001; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR001769; Peptidase C25.
DR InterPro; IPR005536; Peptidase_C25_C.
DR Pfam; PF01364; Peptidase_C25; 1.
DR Pfam; PF03785; Peptidase_C25_C; 1.
KW Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT SIGNAL 1 24
FT PROPEP 25 227
FT CHAIN 228 991 GINGIPAIN R1
FT ACT_SITE 438 438 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 471 471 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 264 265 RT -> TK (IN REF. 2).
SQ SEQUENCE 991 AA; 108782 MW; 03EE3F43CEBE2544 CRC64;

```

QY 382 RDGKIKKEGLTETTFEDGVATGNHXYCVKVBKVTAGVSPKVCVN 425
 DB 872 RDGKIKKEGLTETTYRDAGNSAQSHCYCVKVBKVTAGVSPKVCVD 915

RESULT 5

ATL STAAU STANDARD; PRT; 1256 AA.
 AC P22081;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bifunctional autolysin precursor [includes: N-acetylmuramoyl-L-alanine
 DE anidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
 DE acetylglucosamidase (EC 3.2.1.96)].
 GN ATL.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
 RC STRAIN=RM450;
 RX MEDLINE=95116542; PubMed=7816834;
 RA Ohida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suginaka H.,
 RA Tomasz A.;
 RT "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
 RT alanine amidase domain and an endo-beta-N-acetylglucosaminidase
 RT domain: cloning, sequence analysis, and characterization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 8325-4;
 RA Foster S.J.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN
 CC HIGH-MANNOSE GLYCOPROTEINS AND GLYCOPROTEINS CONTAINING THE
 CC -((MAN)5(GLCNAC)2)-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
 CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
 CC OLIGOSACCHARIDE IS RELEASED INTACT.
 CC -!- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
 CC residues and L-amino acid residues in certain bacterial cell-wall
 CC glycopeptides.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
 CC unit in high-mannose glycopeptides and glycoproteins containing
 CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
 CC residue remains attached to the protein; the rest of the
 CC oligosaccharide is released intact.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
 CC EXTRACELLULAR LYTIC ENZYMES.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
 CC ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
 CC GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D17366; BAA04185.1; -
 CC EMBL; L41499; AAA99982.1; -
 CC InterPro; IPR002502; Amidase_2.
 CC InPro; IPR002901; Amidase_4.
 CC Pfam; PF01510; Amidase_2; 1.
 CC Pfam; PF01832; Amidase_4; 1.
 CC SMART; SM00644; Ami_2; 1.
 CC SMART; SM00047; LYZZ; 1.
 CC Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
 KW SIGNAL
 FT 1 29 POTENTIAL.

FT CHAIN 30 1256 BIFUNCTIONAL AUTOLYSIN.
 FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
 FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
 FT REPEAT 425 589 1.
 FT REPEAT 596 758 2.
 FT REPEAT 770 932 3.
 SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;
 Query Match 5.5%; Score 132.5; DB 1; Length 1256;
 Best Local Similarity 21.9%; Pred. No. 0.15;
 Matches 103; Conservative 47; Mismatches 167; Indels 153; Gaps 25;
 QY 3 PNEPNPNPGTTT---LSESPENGIPASWKITADGDG---NNWTTTTPPP---46
 DB 594 PTPPTPKPSTPTNNKLTVSSINGV---AQINAKNGLFTTVYDKTKPTKEVQKTEA 649
 QY 47 -----GGTSP---AGHNSAICV-----SSASYINPEGPON-PDNYLYTPELSL-----85
 DB 650 VTKEASLGKFKYLVKDYNSPTLIGWVKQGDVYNNNAKSPVNMQTYTVKPGTKLYSPVW 709
 QY 86 -----PGGTTLTFWVCAQDANYASEHYAVYASSTGNDASNFAN-----ALLE 127
 DB 710 GTYKQEAGVSGTGNQTFKATKQOQIDKS---IYLFGTVNGSKSGVSKAYLAVPAAPKK 765
 QY 128 EVLTAKT-----VVTAPPAIR-----GTRVQGTWYQKTVOLPAGTKYVAPRHFG 171
 DB 766 AVAQPNTKAVKVTVPQTQTQTSKIAQVKPNNTGIRASVYEKTAK--NGAKY-ADRTFY 822
 QY 172 CT-----DFFWINLDEVELKANGKEADFE--TPESSTHGEAPAE 209
 DB 823 VTKEAHGNETYVLLNNTSHNIPLGWNVKDLNVQNLGKEVKTQKYTVNKSNGLSMVP 882
 QY 210 WTTIDADGDGQGLCLSSGOLDMLTAHGGTNVVASFSWN-----GMAINPDNLYLSKDV 263
 DB 883 WGT-----KNQVILTGNNI-----AQGFNATKQVSGKDVLYGTINNTGWNNAKOL 931
 QY 264 TGATKVK-----YYVAVNDGPGDHVAVMISKTGT-----NAGDFTVVFEETPN 307
 DB 932 TAPTAVKPTTSAAKDYNVTVIKNG-NGYVVYTPNSDTAKYSLKAFNEQPFVAVKEQVIN 990
 QY 308 G-----INKGARFGLSTEADGAK-----PQSVWIERTVOLPAGTKY 344
 DB 991 GQWYVYKGLSNGKLAWIKST--DLAKELIKYNOTGMTLNQVAQIQAQLQY 1038
 RESULT 6
 GUXB CELFI STANDARD; PRT; 1090 AA.
 AC P50899;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellobiohydrolase B)
 DE (1,4-beta-cellobiohydrolase B) (CBP120).
 GN CBHB OR GENE.
 OS Cellulomonas fimi.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
 RC STRAIN=ATCC 484;
 RX MEDLINE=96003898; PubMed=7575482;
 RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RA "Cellulobiohydrolase B, a second exo-cellobiohydrolase from the
 RT cellulolytic bacterium Cellulomonas fimi.";
 RL Biochem. J. 311:67-74 (1995).
 RN [2]
 RP SEQUENCE OF 54-75.
 RX MEDLINE=93209933; PubMed=8458833;
 RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 RT D (Cend), a family A beta-1,4-glucanase.";


```

J. Bacteriol. 175:1910-1918(1993).
[3]
RN SEQUENCE OF 54-78.
RP MEDLINE=94197708; PubMed=8147863;
RA Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
RT "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
RL CentB, a member of a new family of beta-1,4-glucanases.";
RA Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC -!- FUNCTION: Hydrolyze cellobiose to a mixture of cellobiose,
CC cellobiose and cellobiose, with only a trace of glucose. It
CC hydrolyzed cellobiose to cellobiose and cellobiose, and
CC cellobiose to cellobiose, but it did not hydrolyze cellobiose.
CC Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
CC with inversion of anomeric configuration.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC domain.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY 1 (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L38827; AAB00822.1; -.
DR PIR; S59077; S59077.
DR HSP; P07986; 1EXG.
DR InterPro; IPR001919; Bac_cellose_bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR PRINTS; P00844; GLYDRLASE48.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR SMART; SM00637; CHD_II; 1.
DR SMART; SM00060; FN3; 3.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
KW Cellulose degradation; Hydrolyase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT PROPEP 34 53
FT CHAIN 54 1090
FT CATALYTIC (BY SIMILARITY).
FT FIBRONECTIN TYPE-III 1.
FT FIBRONECTIN TYPE-III 2.
FT FIBRONECTIN TYPE-III 3.
FT CELLULOSE-BINDING (BY SIMILARITY).
FT NUCLEOPHILE (BY SIMILARITY).
FT BY SIMILARITY.
FT DISULFID 990 1089
FT SEQUENCE 1090 AA; 114829 MW; 046BB9D956P2F399 CRC64;
Query Match 5.1%; Score 122.5; DB 1; Length 1090;
Best Local Similarity 21.9%; Pred. No. 0.61;
Matches 102; Conservative 47; Mismatches 157; Indels 159; Gaps 25;
QY 26 ASWKT---IDAGDGNWTTTPPGSTFAGHNSAICVSSAYINPEGPQNDVLYVTEPE 82
DB 515 ASWKPSELKWTGKPTWNAAPTG-----NPLTVEVTYS----- 550
QY 83 LSLPGGGTUTFWVCAQDANYASE---HYAVYASSTGNDAS--NFANALLREV-----LT 131
DB 551 -----GQGVGADTARALLFYAAKSGDTASRKAKALLDAIWANNQDPLG 596

```

```

132 AKTVVT-----APEAIRGTRVQG-----TWYQKTVQLP 159
597 VSAVETRGDYKRFDDTVVANGDGIYIPSGWTGTGTPNGDVIKPGVSVFLDIRSFYKKD---P 653
160 AGTKYVAFRHRGCTDFWINDVEIKANGKRAKRAFTETFESESTGEAPAEWTTIDADGG 219
654 NWSKVQTFELDGGABPQFRYHRFQAQTAVAGALADYARLFDDGT-----TTPDTTAP- 704
220 QGWLCLSSGQLDMLTAHGTTNNVASFSNMGALPNPNVLSKDV--TGATKVKYIYAVNDG 278
705 ----TVPIFG-----LQAGVVTSTETATISWT--ASTDDTRVTGYDVRGATKV----- 745
279 FPGDHYAVMISKTGTNAGD---FTVVFETPNKNGKGAFFGLSTEA---DGAKEPQSVMI 332
746 --GTATTISFTDTGLTASTAYATVTRAFDAAGNVSPSAALTIVTKATPSDTTAP----- 798
333 ERTVDLPAGTKYVAFRHYNCSDLNVLDDLTQFMGSGFTPTDYT-----YTVYRGT 385
799 ----SVPAITS-----SSSTANSV-----TIGWSAS--TDNAGSGSLAGYDVRGAT 839
386 KIKEGLTETTFEEDGV--ATGNHEYCVKVTAG--VSPKVCVNVVT 427
840 RVAQ--TTLATFTDTGLTASTAYEYTVTRADVAGNVSAFSTAVSVT 883
RESULT 7
ID YEEJ ECO57 STANDARD; PRT; 2660 AA.
AC Q8X8V7; Q8X2B9; Q8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein YeeJ.
GN Z3135 OR ECS2775/ECS2775.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI TaxID=83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch).

```

CC CC ENBL; AE005423; AAG57041.1; -
DR ENBL; AP002559; BAB36198.1; ALT_FRAME.
DR ENBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003353; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR000601; PKD.
DR Pfam; PF02369; Big_1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00089; PKD; 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834 BIG-1 1.
FT DOMAIN 840 929 BIG-1 2.
FT DOMAIN 931 1033 BIG-1 3.
FT DOMAIN 1042 1132 BIG-1 4.
FT DOMAIN 1134 1236 BIG-1 5.
FT DOMAIN 1245 1335 BIG-1 6.
FT DOMAIN 1337 1439 BIG-1 7.
FT DOMAIN 1448 1539 BIG-1 8.
FT DOMAIN 1548 1652 BIG-1 9.
FT DOMAIN 1653 1750 BIG-1 10.
FT DOMAIN 1751 1855 BIG-1 11.
FT DOMAIN 1856 1957 BIG-1 12.
FT DOMAIN 1963 2056 BIG-1 13.
FT DOMAIN 2065 2156 BIG-1 14.
FT DOMAIN 2157 2252 BIG-1 15.
FT DOMAIN 2254 2355 BIG-1 16.
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match
Best Local Similarity 5.0%; Score 120.5; DB 1; Length 2660;
Matches 105; Conservative 64; Mismatches 192; Indels 111; Gaps 26;

QY 13 TTTLSESEFE---NGIPASWKTIDAGDGNWTTTPPGGTSFAGHNAICVSSASYNFE 69
DB 1470 TATVKDQDFNEVNNLPVTFSTA-----SSGLTLTPGESNT---NESGIAQATLAGVAF- 1519

QY 70 GPQNPVNYLTPPELSLPGG---TLTF---WVCQADANYASEHVAVVASTGNDASFPAN 123
DB 1520 GEQT-----VTASLANNASDNKTVHFGDTAAAKIIELTPVPSIIAGTPQNSGS--- 1571

QY 124 ALLEBVLTAKTVVTAPPAIRTRVQGTWYQKTVQVLPAGTKYVAFRHFQCTDFFWINLDEV 183
DB 1572 -----VITATVVDNNGPPVKGVTVNFTSNAATAEMTNGGQAVTNEQKATVYTNTRSSI 1626

QY 184 EIKANGKRADETFETFESETHGEAPAEWTTIDADGQO-GMCLSSGQLDMLTAHGGTNV- 241
DB 1627 E---SGARPD---TVEASLENGSSTLSTSNVNDASTAHLTLQLALFDTVSAGDTTNLY 1680

QY 242 --VASFSWNG-----MALNPVNLISKDVTKVKYVAVNDGPPGDHYAVWISKTG 292
DB 1681 IEVKDNYGVGPQOEVLVSVP-----SEGVTPSNALY-----TTNHDGNFYA---SFTA 1728

QY 293 TNAGDFTVW-----FEETPNGI-NKGARFGLSTEADGAKPQSVMIERTVDLPAGTK 343
DB 1729 TKAGVYQVATLENGDSNQQTIVVPVNVANAEISLAASKDPV-----IANNNDLTTLTA 1782

QY 344 YVAFRHYNC-----SDNLILLDDIQ--FTMGSPPTEDTYTVYVYRDGKI---KEGLTETT 395
DB 1783 TVADTEGNAIANSEVTFPLPDEVANFTLG-----DGGKVDWTDTEGRAKVT 1828

QY 396 FEEDGVATGNHEYCEVVKYTAGVSPKVCVN-----VTINPTQFNPCVN 438
DB 1829 LK--GTRAGAH--TVTASMACGSEQLVNFVFIADTLTAQVNLNVTEDNFIAN 1876

```

RESULT 8

PME_ASPAC

ID_PME_ASPAC

AC Q12535;

STANDARD;

PRT; 331 AA.

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pectinesterase precursor (EC 3.1.1.11) (Pectin methyltransferase) (PE).
GN PME1.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KSM 510;
RX MEDLINE=97079238; PubMed=8920970;
RA Christgau S., Kofod L.V., Halkier T., Andersen L.N., Hockauf M.,
RA Dorreich K., Dalboege H., Kauppinen S.;
RT "Pectin methyl esterases from Aspergillus aculeatus: expression
RT cloning in yeast and characterization of the recombinant enzyme.";
RL Biochem. J. 319:705-712(1996).
CC -|- FUNCTION: Involved in maceration and soft-rotting of plant tissue.
CC -|- CATALYTIC ACTIVITY: Pectin + N H(2)O = N methanol + pectate.
CC -|- SIMILARITY: Belongs to the pectinesterase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U49378; AAB42153.1; -.
CC InterPro; IPR000070; Pectinesterase.
CC Pfam; PF01095; Pectinesterase; 1.
CC PROSITE; PS00800; PECTINESTERASE_1; 1.
CC PROSITE; PS00503; PECTINESTERASE_2; 1.
KW Hydrolase; Aspartyl esterase; Cell wall; Signal.
FT SIGNAL 1 17
FT CHAIN 18 331 BY SIMILARITY.
FT ACT_SITE 162 162 BY SIMILARITY.
FT ACT_SITE 183 183 BY SIMILARITY.
SQ SEQUENCE 331 AA; 35681 MW; 1F1C81BF1E32174F CRC64;

Query Match
Best Local Similarity 4.9%; Score 117.5; DB 1; Length 331;
Matches 70; Conservative 33; Mismatches 98; Indels 105; Gaps 15;

QY 108 AVVASSTGNDASNFANALLEVLTAKTVVTAPPAIRTRVQGTWYQKTVQVLPAGTK-KYVA 166
DB 27 AIVVAKSGDYTTIGDAI--DALSTSTDTQTIFIE---EGT-YDEQVYLPANTGVII 79

QY 167 FRHFGCTDFFWINL-----DEVEIKANGKRADETFE-----SSTHGEAPAE 209
DB 80 YGQTEINTDSYADNLVITTHAISYEDAGESDDLATFRNKAVGQVYVNLNANTCGQACHQ 139

QY 210 WTTIDADGQOQWL-CLSSGQLDMLTAHGGTNV----- 242
DB 140 ALALSADAWDQGYGNCFTGYQDTLLAQTNQNLVINSYIEGAVDFTFGQHARAWFQNVDI 199

QY 243 -----ASFWSNMGALNPDN--YLISKDVTGA-----TKVKYVAVNDGPPGDHYAV 286
DB 200 RVVEGPTASITANGSRSETDTSYVINKSTVAAKESDDVAEGTYIL----GPPWSEYA- 254

QY 287 MISKTGTNAGDFTVVFEEET--PNGINKGGARFGLSTEADGAKPQSVMIERTVDLPAGTKY 344
DB 255 -----RVVFQOTSMTNVNSLG-----WTEWSTSTP-NTEY 284

QY 345 VAFRHY 350
DB 285 VTFGEY 290

```

RESULT 9

YK28_ARCFU

```

ID YK28_ARCFU STANDARD; PRT; 607 AA.
AC O28251;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2028.
GN AF2028.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.H., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artisch P., Kaine B.P., Sykes S.M.,
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO M.JANNASCHII M31393 AND M31394.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000963; AAB89227.1; -.
CC TIGR; AF2028; -.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 16 36 POTENTIAL.
CC TRANSMEM 583 603 POTENTIAL.
CC SEQUENCE 607 AA; 66215 MW; F46F116BF2AAFE1 CRC64;
Query Match 4.9%; Score 117.5; DB 1; Length 607;
Best Local Similarity 18.3%; Pred. No. 0.64;
Matches 86; Conservative 65; Mismatches 172; Indels 147; Gaps 23;
QY 9 PNPQTTLSESPENGIPASMTID-ADG-----DGNWTTTPPGGTSFAGHN 55
DB 201 PSSVTLTKNNLNTSYTNWKTLEFADGSSNPQIISGLTNNKWLTT----- 249
QY 56 SAICVSSASYINFGFQPNPNLVLPFELSPGGGTLTFWCAQDAN-----YASEH 106
DB 250 -----FKVKNYSTDEAFVS---RLVGFGEAYFQFGPINGNSGSKIIDVFAIGN 296
QY 107 YAVYASSTGNDSN--FANALLEEVLTAKTVVTAPEARTGT-----RVQGTWYQTVLP 159
DB 297 ASIGVKNKSPNSQNTGNVTIKNTATGLTIVKSVKVMATDRNYNEINGARYENT--- 353
QY 160 AGTKYVAFRHFGCTDFFWNLNDEVIKANGRADFTETFESTHGEAPEWTTID---AD 216
DB 354 -----VN-----VQI-----GRDESFTSKDLSFYDKVPLINGNVTFLRVE 389
QY 217 GDGQWGLCLSSGQDLTAHGCTNTVAVSFWNGMALNPDPNLYISKDVTKATVKYVAVN 276
DB 390 DANYGW---GVQQ-DKIIDGGNTYIIERI-----YVIGSVLYKVTK-HVESAGN 433
QY 277 DGFPGDGVAVMISKTGTNAGDFTVTVFEETPNKGGARFGLST-EADGAKPQ-----SWVI 332
```

```

DB 434 DIY---NITLWENLGQESFYVYVDLIPK-----NFSLTNGDNDWKDPQDRGDWV 483
QY 333 ERTVDLPAGTKYVAFRHYNGSDLNVLDDDIQFTMGGSPTPTDYTYV----- 380
DB 484 NKSSMLAGGPETIT-----NIQL-----SGYDTGWWRIRPINASADGDGA 524
QY 381 YRDGTKIKEGLTETTFEEDGVATGNHVEYCVVEKYTAGVSPKVCVNVITNP 430
DB 525 YDDVTEIENNTQTVVIFYQ---IQGSDEYKLLDAIVGIDPILSMNEQTSP 571
RESULT 10
ADP1_MYCGA STANDARD; PRT; 1122 AA.
ID ADP1_MYCGA
AC Q49379; Q49437; Q53351;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN GAPA OR MGCI.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S6;
RX MEDLINE=96201559; PubMed=8613358;
RA Keeler C.L. Jr., Hnatow L.L., Whetzel P.L., Dohms J.E.;
RT "Cloning and characterization of a putative cytoadhesin gene (mgcl)
RT from Mycoplasma gallisepticum.";
RL Infect. Immun. 64:1541-1547(1996).
RN [2]
RP SEQUENCE OF 1-12 FROM N.A.
RC STRAIN=S6;
RX Hnatow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 378-570 FROM N.A.
RC STRAIN=S6;
RX MEDLINE=93371270; PubMed=8363503;
RA Dohms J.E., Hnatow L.L., Whetzel P., Morgan R., Keeler C.L. Jr.;
RT Identification of the putative cytoadhesin gene of Mycoplasma
RT gallisepticum and its use as a DNA probe.";
RL Avian Dis. 37:380-388(1993).
RN [4]
RP SEQUENCE OF 159-1122 FROM N.A.
RC STRAIN=S6;
RA Goh M.S., Geary S.J.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Could be involved in cytoadherence.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the adhesin P1 family.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC frameshift in position 159 to 213.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U34842; AAB02987.1; -.
CC EMBL; U44804; AAC83385.1; ALT_FRAME.
CC FIR; I18346; I18346.
KW Cytoadherence; Signal; Transmembrane.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1122 ADHESIN P1.
FT TRANSMEM 997 1021 POTENTIAL.
FT DOMAIN 1002 1009 POLY-ILE.
FT CONFLICT 313 314 DM -> IW (IN REF. 4).
```

```

FT CONFLICT 382 382 Y -> F (IN REF. 3).
FT CONFLICT 501 501 I -> T (IN REF. 3).
FT CONFLICT 568 568 R -> G (IN REF. 3).
FT CONFLICT 570 570 T -> A (IN REF. 3).
FT CONFLICT 693 695 DIL -> VYT (IN REF. 4).
SQ SEQUENCE 1122 AA; 121305 MW; 155C34DA2D6C3C65 CRC64;

Query Match 4.9%; Score 117; DB 1; Length 1122;
Best Local Similarity 19.5%; Pred. No. 1.6; Indels 195; Gaps 24;
Matches 93; Conservative 78; Mismatches 195; Indels 112; Gaps 24;

QY 42 TTPPGGTSPAGHNSAICVSSASYINFEQPQN-----PDNVLVTPELSLPGGGT 90
DB 437 TAEAPGNTKTVGPYGLLSA---ISFDATRNLALAPALQGDVGHYHFLAVGVSS 493
QY 91 LTFWVCADANYASEHVAVASSTGNDASNFAN-----ALLBEVLTAKTVTVAPEAIRG 144
DB 494 -----PRGAN--GNIFLGSAITWGTNGNGLDTPKWHSPAVIEDAPTFITVNSSGVLQN 545
QY 145 TRVQGTWYQKTVQLP--AGTKYVAFRHFGCTDFFWINLDEVEIK---ANGKRAD--FTET 197
DB 546 S---GSQOSTPMPNSNGNESIPYRTNSVDYNSVRFALISKPAGNTKQVESLFTTA 602
QY 198 FESSTHGEAPAEWTID-----ADGGQGWCLSSGQLDNLTAHGNTNVVAFSFWN--- 248
DB 603 LKLDTLNLSLNFKEQNNIFFSYAMLDRGOWSLGTRKDSWLT-----TNTINNFTYNTQQ 658
QY 249 -----GMALNPNDYLIS-----KDV-----TGATKVKYVAVND----- 277
DB 659 QLASTAGENANPRNINLALTAKGPDRIIDGNVYTSNNTKFTYYQVGGAITTWE 718
QY 278 ---GFGPDHYAVMLSKTGTNAGDFTVVFEETPNGINK--GGARFGLSTEADGAKPQSVWIE 333
DB 719 VQVNYKTSANITYNLTTRTFDGTTPATQDANTVSSKINGAYLSSTGQGWYNGSIYVK 778
QY 334 RTVDLPAGTKV--AFRHYNCSDNLNILLDDIOFTMGSGSPPTDITYTVVRDGTKEGL 391
DB 779 KASFTPSQGYTWQDFKGLTTASNAVISN---WTKAG-----YSIRPDDDTV-FNV 826
QY 392 TETTFEEDGVATGN---HEYCVB-----VKYTAGVSPKVCVNVITNPTQF-NPVON 438
DB 827 SKIPFEKEITAAVNRSLDSYVQLNGETSVNTVARVSPDSSA-LTLNPKRITNPLMN 883

RESULT 11
ADP2 MYCGA
ID ADP2 MYCGA STANDARD; PRT; 1122 AA.
AC Q9REB8; 2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein) (Adherence protein A).
DE GAPA OR MYCGA1800 OR MGA_0934.
GN Mycoplasma gallisepticum.
OS Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R(low);
RX MEDLINE=22830409; PubMed=12949158;
RA Papazisi L., Gorton T.S., Kutish G., Markham P.F., Browning G.F.,
RA Nguyen D.K., Swartzell S., Madan A., Mahairas G., Geary S.J.;
RA "The complete genome sequence of the avian pathogen Mycoplasma
RT gallisepticum strain R(low).";
RL Microbiology 149:2307-2316(2003).
RN [2]
RP SEQUENCE OF 94-1122 FROM N.A.
RC STRAIN=R(low);
RX MEDLINE=20536401; PubMed=11083776;
RA Papazisi L., Troy K.E., Gorton T.S., Liao X., Geary S.J.;
RA "Analysis of cytoadherence-deficient, GAPA-negative Mycoplasma
RT gallisepticum strain R.;"

```

```

RL Infect. Immun. 68:6643-6649(2000).
CC -I- FUNCTION: Could be involved in cytoadherence (By similarity).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -I- SIMILARITY: Belongs to the adhesin P1 family.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to a license@sib-sib.ch).
CC -----
DR EMBL; AE016967; AAP56530.1; ALT_INIT.
DR EMBL; AF214004; AAP25381.1; -.
KW Cytadherence; Signal; Transmembrane; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1122 ADHESIN P1.
FT TRANSMEM 1001 1021 POTENTIAL.
FT DOMAIN 1002 1009 POLY-ILE.
FT CONFLICT 337 337 F -> C (IN REF. 2).
FT CONFLICT 645 645 T -> A (IN REF. 2).
FT CONFLICT 826 826 V -> A (IN REF. 2).
SQ SEQUENCE 1122 AA; 121387 MW; E57D27A7FDD360AD CRC64;

Query Match 4.7%; Score 113; DB 1; Length 1122;
Best Local Similarity 19.5%; Pred. No. 3;
Matches 93; Conservative 77; Mismatches 196; Indels 112; Gaps 24;

QY 42 TTPPGGTSPAGHNSAICVSSASYINFEQPQN-----PDNVLVTPELSLPGGGT 90
DB 437 TAEAPGNTKTVGPYGLLSA---ISFDATRNLALAPALQGDVGHYHFLAVGVSS 493
QY 91 LTFWVCADANYASEHVAVASSTGNDASNFAN-----ALLBEVLTAKTVTVAPEAIRG 144
DB 494 -----PRGAN--GNIFLGSAITWGTNGNGLDTPKWHSPAVIEDAPTFITVNSSGVLQN 545
QY 145 TRVQGTWYQKTVQLP--AGTKYVAFRHFGCTDFFWINLDEVEIK---ANGKRAD--FTET 197
DB 546 S---GNPQSTPMPNSNGNESIPYRTNSVDYNSVRFALISKPAGNTKQVESLFTTA 602
QY 198 FESSTHGEAPAEWTID-----ADGGQGWCLSSGQLDNLTAHGNTNVVAFSFWN--- 248
DB 603 LKLDTLNLSLNFKEQNNIFFSYAMLDRGOWSLGTRKDSWLT-----TNTINNFTYNTQQ 658
QY 249 -----GMALNPNDYLIS-----KDV-----TGATKVKYVAVND----- 277
DB 659 QLASTAGENANPRNINLALTAKGPDRIIDGNVYTSNNTKFTYYQVGGAITTWE 718
QY 278 ---GFGPDHYAVMLSKTGTNAGDFTVVFEETPNGINK--GGARFGLSTEADGAKPQSVWIE 333
DB 719 VQVNYKTSANITYNLTTRTFDGTTPATQDANTVSSKINGAYLSSTGQGWYNGSIYVK 778
QY 334 RTVDLPAGTKV--AFRHYNCSDNLNILLDDIOFTMGSGSPPTDITYTVVRDGTKEGL 391
DB 779 KASFTPSQGYTWQDFKGLTTASNAVISN---WTKAG-----YSIRPDDDTV-FNV 826
QY 392 TETTFEEDGVATGN---HEYCVB-----VKYTAGVSPKVCVNVITNPTQF-NPVON 438
DB 827 SKIPFEKEITAAVNRSLDSYVQLNGETSVNTVARVSPDSSA-LALNPKRITNPLMN 883

RESULT 12
ALS1 CANAL
ID ALS1 CANAL STANDARD; PRT; 1260 AA.
AC P46590; 1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Agglutinin-like protein 1 precursor.
GN ALS1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

Saccharomycetales; mitosporic Saccharomycetales; Candida.

NCBI_TaxID=5476;

[1]

SEQUENCE FROM N.A. / B792;

STRAIN=ATCC 11651; PubMed=7752895;

MEDLINE=95272392; PubMed=7752895;

Hoyer L.L., Scherer S., Shatzman A.R., Iivi G.P.;

"Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual agglutinin separated by a repeating motif.";

Mol. Microbiol. 15:39-54(1995).

CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.

CC -!- PFM: N-glycosylated and O-glycosylated (Potential).

CC -!- SIMILARITY: TO YEAST SAG1.

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

EMBL: L25902; AAC41649.2; -

InterPro; IPR008440; Candida; ALS.

DR Pfam; PF05792; Candida_ALS; 1.

KW Cell adhesion; Glycoprotein; Repeat; Signal.

FT SIGNAL 1 17

FT CHAIN 18 1260

FT DOMAIN 433 792

FT REPEAT 433 468

FT REPEAT 469 504

FT REPEAT 505 540

FT REPEAT 541 576

FT REPEAT 577 612

FT REPEAT 613 648

FT REPEAT 649 684

FT REPEAT 685 720

FT REPEAT 721 756

FT REPEAT 757 792

FT REPEAT 793 1152

FT REPEAT 983 1043

FT REPEAT 1092 1152

FT DOMAIN 399 404

FT DOMAIN 408 418

FT DOMAIN 450 455

FT DOMAIN 486 491

FT DOMAIN 522 527

FT DOMAIN 558 563

FT DOMAIN 594 599

FT DOMAIN 630 635

FT DOMAIN 666 671

FT DOMAIN 702 707

FT DOMAIN 738 743

FT DOMAIN 774 779

FT DOMAIN 874 877

FT CARBOHYD 471 471

FT CARBOHYD 579 579

FT CARBOHYD 615 615

FT CARBOHYD 687 687

FT CARBOHYD 723 723

FT CARBOHYD 820 820

FT CARBOHYD 886 886

FT CARBOHYD 918 918

FT CARBOHYD 973 973

FT CARBOHYD 1045 1045

FT CARBOHYD 1068 1068

FT CARBOHYD 1260 1260

FT CARBOHYD 132641 132641

FT CARBOHYD 763D1063R2354C24 CRC64;

SEQUENCE 1260 AA; 132641 MW; 763D1063R2354C24 CRC64;

Query Match 4.7%; Score 112.5; DB 1; Length 1260;

Best Local Similarity 19.5%; Pred. No. 3.8;

Matches 68; Conservative 30; Mismatches 77; Indels 173; Gaps 16;

5 PNPNPQTTTL-SESFENGIPASWKTIDAGDGN-----W----- 40

Db 432 PNPNTVSTTTEYSQSF-----ATTTTAPGGTDTTIIREPNHNTVTTTEYSQSFA 486

Qy 41 -TTTPPGGTSFAGHNSAICVSSASYINFGPONP-----DNLVTPPELSLPGGG 89

Db 487 TTTTAPPGGT-----DSVI-----IREPNFTVTTTEYSQSFAITTTTAPPGG 532

Qy 90 TLTFWV-----CAQDANYASEHYAVASSTGNDASNFANALLBEVLFAKTVVTAPEAIR 143

Db 533 TDSVIREPNFTVTTTEYSQSYA-----TTTTTAP----- 565

Qy 144 GTRVQGTWYQKTVQLPAGTKYVAFR-----HFGCTDFFWNLDELVEIKANGKRAADFTETFE 199

Db 566 -----PGTDSVIREPNHNTVTTTEYW-----SQSFA 593

Qy 200 SSTHGEAPAEWTIDADGGQGLCLSSGQLDMLTAHGTTNVASFSWNGMALNPNDNYLI 259

Db 594 TTTTAPPGT-----GTTDTV-----IIREPNHNTV 618

Qy 260 SKDVTGATKVKYKYYAVNDGFGDHYAVNMISKTGNAGDFTVVFEETPN 307

Db 619 -----TTTEYW-----SQSFATTTTGTGPGSGTDTTIIREPPN 651

RESULT 13

GUXA_CELFI STANDARD; PRT; 872 AA.

AC P50401;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Exoglucanase A precursor (Ec 3.2.1.91) (Exocellobiohydrolase A)

DE (1,4-beta-cellobiohydrolase A) (CBP95).

GN CBHA.

OS Cellulomonas fimi.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Micrococcineae; Cellulomonadaceae; Cellulomonas.

OX NCBI_TaxID=1708;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 484;

RX MEDLINE=94344030; PubMed=8065260;

RA Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,

RA Miller R.C. Jr.;

RT "Cellulobiohydrolase A (Cbha) from the cellulolytic bacterium

RT Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to

RT Trichoderma reesei CBH II.";

RL Mol. Microbiol. 12:413-422(1994).

RN [2]

RP SEQUENCE OF 41-58.

RX MEDLINE=93209933; PubMed=8458833;

RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase

RT D (CenD), a family A beta-1,4-glucanase.";

RL J. Bacteriol. 175:1910-1918(1993)

CC -!- FUNCTION: This enzyme hydrolyzes 1,4-beta-D-glucosidic linkages of

CC cellulose. Weak activity against carboxymethylcellulose, bacterial

CC microcrystalline cellulose and barley beta-glucan. Has also weak

CC endoglucanase activity. Hydrolyzes glucosidic bonds with inversion

CC of anomeric configuration.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages

CC in cellulose and cellotetraose, releasing cellobiose from the non-

CC reducing ends of the chains.

CC -!- SIMILARITY: Contains 3 fibronectin type III domains.

CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)

CC domain.

CC -!- SIMILARITY: Belongs to cellulase family B (family 6 of glycosyl

CC hydrolases).

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L25809; AAC36898.1; --
DR PIR; S49541; S49541.
DR HSP; P07986; IEXG.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellul_bnd.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_subd.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00014; ENTPHIII.
DR PRINTS; PR00733; GLHYDRIASE6.
DR ProDom; PD003733; Glyco_hydro_6; 2.
DR SMART; SM00637; CBD_II; 1.
DR SMART; SM00060; FN3; 3.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 40
FT CHAIN 41 872 EXOGLUCANASE A.
FT DOMAIN 41 477 CATALYTIC.
FT DOMAIN 478 563 FIBRONECTIN TYPE-III 1.
FT DOMAIN 573 664 FIBRONECTIN TYPE-III 2.
FT DOMAIN 673 768 FIBRONECTIN TYPE-III 3.
FT DOMAIN 769 872 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 188 188 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 410 410 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 140 202 BY SIMILARITY.
FT DISULFID 374 428 BY SIMILARITY.
FT DISULFID 770 869 BY SIMILARITY.
SQ SEQUENCE 872 AA; 89300 MW; 7883B407F995533B CRC64;

Query Match 4.7%; Score 112; DB 1; Length 872;
Best Local Similarity 24.0%; Pred. No. 2.5;
Matches 85; Conservative 28; Mismatches 113; Indels 128; Gaps 21;

Qy 12 GTTTLSSFFENGIPASWKTDADGDNWTTTPPGGTSPAGHNSAICVSSASYINPEGP 71
Db 586 GTTTT-----VPLSWT-----ASTDNAGSGVAGYE---VLRGTTVV---GT 623

Qy 72 QNPDNVLTPELSLPGGGTLTFWYCAQD--ANYASEHVAVYASS-----TG 115
Db 624 TTATSYT-----GLTAGTYSFSVRKDVAGNTSAASAASAITQTGTVDVTTAPSPTG 680

Qy 116 NDASNANALAEVLTAKT-----VVTAPAIRGT-----RVQGTWYOKTVQLPAGTKY- 164
Db 681 LTAGTTTSSVPLTWASTDNAGSGVAGYEVFNGTTRVATVTSYTVT--GLAADTAYS 739

Qy 165 -----VAFRHFGCTDFWNLNDEIVEIKANGKRADFTETPESS 201
Db 740 FTVKAKDVAGNVSAASAASARTQAATSGGCT-----VKYSASSWNTGFTGTVEVK 790

Qy 202 THGRAPAEWTTID--ADGD--GGWLCSSGQDLTAHGGTNVVA--SFSWNGMALNPD 255
Db 791 NNGTAALNGWTLGRSFDGQKVSQW-----SAEW--SQSGTAVTAKNAPWNTGL----- 838

Qy 256 NYLISKDVGTATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVFPEETNGI 309
Db 839 -----AAGS-----SVSIGENGTH-----NGIN-----TAPTAFTLNGV 867

RESULT 14
ID SLAP LACBR
AC Q05044;
DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE S-layer protein precursor (Surface layer protein).
OS Lactobacillus brevis.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1580;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=ATCC 8287 / DSM 20556;
RX MEDLINE=93054358; PubMed=1429463;
RA Vidgren G., Palva I., Pakkanen R., Lounatmaa K., Palva A.;
RT "S-layer protein gene of Lactobacillus brevis: cloning by polymerase
chain reaction and determination of the nucleotide sequence.";
RL J. Bacteriol. 174:7419-7427(1992).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a
S-layer with tetragonal symmetry.
CC -!- SIMILARITY: SOME, TO THE S-LAYER PROTEIN OF L.ACIDOPHILUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z14250; CAA78618.1; --
DR PIR; A47023; A47023.
KW Signal; Glycoprotein; Cell wall; S-layer.
FT SIGNAL 1 30
FT CHAIN 31 465 S-LAYER PROTEIN.
SQ SEQUENCE 465 AA; 48159 MW; 2BE240392E65A2C CRC64;

Query Match 4.6%; Score 111; DB 1; Length 465;
Best Local Similarity 19.4%; Pred. No. 1.3;
Matches 85; Conservative 45; Mismatches 197; Indels 122; Gaps 16;

Qy 13 TTTLSSEFEN-GIPASWKTDIDA---DGDGNWTTTPPGGTSPAGHNSAICVSSASYIN 67
Db 26 TTASAKSYATAGAYSTLIKIDAAFNVEATGTNLYTKP--GTV--KGAKVASKATMAK 80

Qy 68 FEGQPNPNVLTPELSLPGGGTLTFWYCAQDANYASEHVAVYASSTGNDASHPANALLE 127
Db 81 LASSKKSADYFRAYGVKVTNRGSVYRVVTMDGKYGVYVG-----GKSDTAPAGGI-- 132

Qy 128 EVLTAKTVVTAPAIRGTRVQGTWYOKTVQ--LPAGTKYVAFRHFGCTDFWNLNDEIVEI 185
Db 133 -----KSAETTTKADMPARTTGFVLTDTSKNTLTWPKYQYK-----ASKVSL 176

Qy 186 KANGKRADFTETPESSTHGEAPAEWTTIDADGDG--QGWLCSSGQDLTAHGGTNVVAS 244
Db 177 YGVAKDTKFTVDQAATKTRGSLYVHTATNGSGISGWIYAGKG---FSTTATGTQVLGG 233

Qy 245 FSNWGMALNPDNLYISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVFEE 304
Db 234 LSTDK-----SVTATNDSVKIVYRTD-----GTQVGSNTWV--- 266

Qy 305 TPNGINKGARFGLSTEADGAKPQSVWIERTVDPAPGKTV-----AFRHNCSDLN 356
Db 267 -----TSTDGTGKAGSKVSKAAAOQTALAEVINANKPSGYVTVPNADAT 311

Qy 357 Y-----ILLDDIQFTMGSPPTDYT-----YTVVRDG 384
Db 312 YGNTVYATVSAATS KVALKVSGETVTTALTADANDKVAANDTTANGSSVAGSTVAAAG 371

Qy 385 TKKEGLTETTFEDGVAT 403
Db 372 TKLAQLTDLTGKGVVVT 390

```
RESULT 15
YD94 METJA
ID YD94 METJA STANDARD; PRT; 987 AA.
AC Q58789;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB Hypothetical protein MJ1394.
GN MJ1394.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RT Science 273:1058-1073 (1996).
RL
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67579; AAB99404.1; -.
DR PIR; A64474; A64474.
DR TIGR; MJ1394; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 958 978
SQ SEQUENCE 987 AA; 112360 MW; DLE628FFB28CA86D CRC64;
Query Match 4.6%; Score 110; DB 1; Length 987;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 81; Conservative 69; Mismatches 135; Indels 124; Gaps 24;
QY 6 NPNPNPGITTLSEFNGI-----PASW--KTIDADG-DGNNTWTTPPGGTSE-AG 53
DB 429 NNDPNNGVKLLPKETGRDVIGWVYRPSNWGGPIDRIGLEENF-----DGYSFEVN 482
QY 54 HNSAICVSSASYINFEQPQNPQNYLVTPELSLPGGGLTFWVCAQDANYASEHYAVYASS 113
DB 483 HVS-----NVISIDRRNGNPTETSPV-----YWNPPDEWYTFE-LKIYSNG 525
QY 114 TGNDAENFANALLEVLTAQTVVTAPEAIRGTRVQGTWYQVQLPAGTKYVAFRFGCT 173
DB 526 TITFTYYQNGSL-----AATVST-----IDNT-YTKFDVVVHGYYVY----- 564
QY 174 DFFWINLDEVEIKANGKRADETETPESSSTHGEAPAEWTTIDADGGQGWLCISSLGQDLW 233
DB 565 -----VDLEV--NSKNDF-----YGNKWKILEITANSSEGTAVLFDG--DYF 606
QY 234 TAAGGTNVVASFPSWNGMALNPQNYLISKDVTGATKY-----KYTAVND-----GFPGD 282
DB 607 KXDYNTSNLAINWTNITLNSN-----DSATLVFNVLGNYSYSERDNLAKYGFA-- 657
QY 293 HYAVMISKGTGNAGDFTVVFEETPNKNGGARFGLSTADGAKPOSVMIERTVLDLPGT 342
```

Db 658 --KILFNYNGTNT-----NTSIKGVYAGS--YSISTDHGTTTGEINWIEN----- 699

QY 343 KYVAFRH-----YNCSDNLNVLDDIQFTWGGSPPTDYYTYTVYRDGT 385

DB 700 --VTFKNDAKSYSPNLNINAIWVKNKSAELYWNP-----FNKSIWIDGS 742

Search completed: May 18, 2004, 11:43:37
Job time : 10.2235 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:36:20 ; Search time 32.094 Seconds
(without alignments)
4315.838 Million cell updates/sec

Title: US-08-570-311-22

Perfect score: 2388

Sequence: 1 GTPNPNPNGTTLSESF.....PKVCNVNINPTQNPVQNL 439

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum Match 0%

Maximum DB seq length: 2000000000

Listing first 45 summaries

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------------|
| 1 | 2074 | 86.9 | 1723 | 2 P72197 | P72197 porphyromon |
| 2 | 2068 | 85.6 | 1723 | 2 P72194 | P72194 porphyromon |
| 3 | 2060 | 86.3 | 1358 | 2 P96967 | P96967 porphyromon |
| 4 | 2058 | 86.2 | 1687 | 2 Q9R9B7 | Q9R9B7 porphyromon |
| 5 | 2058 | 86.2 | 1704 | 2 Q51816 | Q51816 porphyromon |
| 6 | 2056.5 | 86.1 | 1223 | 2 Q29NB5 | Q29NB5 porphyromon |
| 7 | 2048 | 85.8 | 1706 | 2 Q51839 | Q51839 porphyromon |
| 8 | 2037 | 85.3 | 1706 | 2 Q51838 | Q51838 porphyromon |
| 9 | 2030.5 | 85.0 | 1097 | 2 P72196 | P72196 porphyromon |
| 10 | 2030.5 | 85.0 | 1732 | 2 O07442 | O07442 porphyromon |
| 11 | 2025.5 | 84.8 | 1732 | 2 Q51817 | Q51817 porphyromon |
| 12 | 2011.5 | 84.2 | 1732 | 2 O52050 | O52050 porphyromon |
| 13 | 691.5 | 29.0 | 925 | 2 Q9F4J0 | Q9F4J0 porphyromon |
| 14 | 571.5 | 23.9 | 312 | 2 Q9KIB3 | Q9KIB3 porphyromon |
| 15 | 199 | 8.3 | 293 | 2 Q9XBU9 | Q9XBU9 porphyromon |
| 16 | 136.5 | 5.7 | 1341 | 16 Q8UAU1 | Q8UAU1 Quaul agrobacteri |

| | | | | | | |
|----|-------|-----|------|----|--------|---------------------|
| 17 | 134.5 | 5.6 | 680 | 2 | Q52644 | Q52644 ruminococcu |
| 18 | 134 | 5.6 | 2656 | 5 | Q9GNU3 | Q9GNU3 paracentrot |
| 19 | 132.5 | 5.5 | 555 | 16 | Q826W1 | Q826W1 streptomycy |
| 20 | 132.5 | 5.5 | 1256 | 16 | Q8NX96 | Q8NX96 staphylococ |
| 21 | 132 | 5.5 | 1742 | 16 | Q8P377 | Q8P377 xanthomonas |
| 22 | 131.5 | 5.5 | 1744 | 16 | Q82YW8 | Q82YW8 enterococcu |
| 23 | 131.5 | 5.5 | 2215 | 16 | Q7WBN0 | Q7WBN0 bordetella |
| 24 | 129.5 | 5.4 | 1248 | 16 | Q99V41 | Q99V41 staphylococ |
| 25 | 129.5 | 5.4 | 1255 | 2 | Q7WTC6 | Q7WTC6 staphylococ |
| 26 | 129.5 | 5.4 | 2768 | 16 | Q8E9G6 | Q8E9G6 shewanella |
| 27 | 129.5 | 5.4 | 3346 | 16 | Q7WNS4 | Q7WNS4 bordetella |
| 28 | 129 | 5.4 | 691 | 16 | Q9RZ57 | Q9RZ57 deinococcu |
| 29 | 128.5 | 5.4 | 2230 | 16 | Q7U7J7 | Q7U7J7 synechococc |
| 30 | 128 | 5.4 | 756 | 9 | Q858B6 | Q858B6 enterobacte |
| 31 | 126 | 5.3 | 2734 | 16 | Q89C73 | Q89C73 bradyrhizob |
| 32 | 124 | 5.2 | 1608 | 17 | Q8PVI0 | Q8PVI0 methanosarc |
| 33 | 124 | 5.2 | 3988 | 17 | Q8TPZ1 | Q8TPZ1 methanosarc |
| 34 | 124 | 5.2 | 5188 | 16 | Q8X4H5 | Q8X4H5 escherichia |
| 35 | 124 | 5.2 | 5291 | 16 | Q8X2T1 | Q8X2T1 escherichia |
| 36 | 124 | 5.2 | 7716 | 16 | Q7UWZ8 | Q7UWZ8 rhodospirill |
| 37 | 123.5 | 5.2 | 1095 | 2 | Q85151 | Q85151 photorhabdu |
| 38 | 123.5 | 5.2 | 1672 | 16 | Q8Y366 | Q8Y366 ralstonia s |
| 39 | 123.5 | 5.2 | 1684 | 2 | Q83658 | Q83658 unidentified |
| 40 | 123 | 5.2 | 452 | 10 | Q8H753 | Q8H753 phytophthor |
| 41 | 123 | 5.2 | 596 | 2 | Q60029 | Q60029 thermomonos |
| 42 | 123 | 5.2 | 1119 | 2 | Q84DC8 | Q84DC8 mycoplasma |
| 43 | 122 | 5.1 | 2522 | 16 | Q8EKA6 | Q8EKA6 shewanella |
| 44 | 122 | 5.1 | 2951 | 17 | Q8THC9 | Q8THC9 methanosarc |
| 45 | 121 | 5.1 | 1649 | 16 | Q9CFA2 | Q9CFA2 lactococcu |

ALIGNMENTS

RESULT 1

P72197 PRELIMINARY; PRT; 1723 AA.

AC P72197; 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Lys-gingipain.

GN KGP.

OS Porphyromonas gingivalis (Bacteroides gingivalis).

OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;

OC Porphyromonadaceae; Porphyromonas.

OX NCBI_TaxID=837;

RN [1]

RP SEQUENCE FROM N.A.

RA Pavloff N., Pemberton P.A., Potempa J., Chen W.-C.A., Pike R.N.,

RA Prochazka V., Kiefer M.C., Travis J., Barr P.J.;

RT "Molecular cloning and characterization of Porphyromonas gingivalis

RT Lys-gingipain."

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; U54691; AAA99810.1; -.

DR MEROPS; C35.002; -.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.

DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.

DR GO; GO:0006310; F:DNA recombination; IEA.

DR GO; GO:0006281; F:DNA repair; IEA.

DR GO; GO:0006260; F:DNA replication; IEA.

DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000977; DNA ligase.

DR InterPro; IPR001769; Peptidase_C25.

DR InterPro; IPR005536; Peptidase_C25_C.

DR Pfam; PF01364; Peptidase_C25; 1.

DR Pfam; PF03785; Peptidase_C25_C; 1.

DR PROSITE; PS00697; DNA_LIGASE_A1; 1.

SQ SEQUENCE 1723 AA; 186831 MW; 4508A7E50197CEBD CRC64;

Query Match 86.9%; Score 2074; DB 2; Length 1723;

Best Local Similarity 89.1%; Pred. No. 2.9e-129;

Matches 392; Conservative 11; Mismatches 31; Indels 6; Gaps 4;

Qy 1 GTPNPNPNPGTTTISESFENGIPASWKTIDADGGNNMTTTPPGGTSPAGHNSAICV 60
 Db 969 GTPNPNPNPGTTTISESFENGIPASWKTIDADGGHGWKPGNAPG---IAGYNSNGCV 1025

Qy 61 SSASY-INFEQPONPDNYLTPPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
 Db 1026 YSEFGLGGIGVLPEDNYLITPALDLPGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy 120 NFANALLEEVLTAKTVVTAPRAIGTRVQGTWYOKTVLPAGTKYVAFRHFQCTDFFWIN 179
 Db 1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy 180 LDEVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQDLWLTAHGGT 239
 Db 1144 LDEVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQDLWLTAHGGT 1203

Qy 240 NVVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGPFQGDHYAVMISKTGTNAGDFT 299
 Db 1204 NVVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGPFQGDHYAVMISKTGTNAGDFT 1263

Qy 300 VVFETPNKNGKARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
 Db 1264 VVFETPNKNGKARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy 360 LDDIQFTMGSGPTPTDYTYTVYRDGTKIKEGLTTFEEDGVATGNHHEYCVVEKYTAGVS 419
 Db 1324 LDDIQFTMGSGPTPTDYTYTVYRDGTKIKEGLTTFEEDGVATGNHHEYCVVEKYTAGVS 1383

Qy 420 PKVCNVNTINPTQFNPNQNL 439
 Db 1384 PKVCNVNTINPTQFNPNQNL 1403

RESULT 2

ID P72194 PRELIMINARY; PRT; 1723 AA.
 AC P72194;
 DT 01-FEB-1997 (TREMELrel. 02, Created)
 DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Lys-gingipain.
 DE KGP.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 GN Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RX MEDLINE=97044756; PubMed=8889827;
 RA Okamoto K., Kadowaki T., Nakayama K., Yamamoto K.;
 RT "Cloning and sequencing of the gene encoding a novel lysine-specific
 RT cysteine proteinase (Lys-gingipain) in Porphyromonas gingivalis:
 RT structural relationship with the arginine-specific cysteine proteinase
 RT (Arg-gingipain).";
 RL J. Biochem. 120:398-406 (1996).
 DR EMBL; D83258; BAAL1870.1; -.
 DR MEROPS; C25.002; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.

SQ SEQUENCE 1723 AA; 187261 MW; 5628963D251493EB CRC64;

Query Match 86.6%; Score 2068; DB 2; Length 1723;
 Best Local Similarity 88.9%; Pred. No. 7.2e-129;
 Matches 391; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

Qy 1 GTPNPNPNPGTTTISESFENGIPASWKTIDADGGNNMTTTPPGGTSPAGHNSAICV 60
 Db 969 GTPNPNPNPGTTTISESFENGIPASWKTIDADGGHGWKPGNAPG---IAGYNSNGCV 1025

Qy 61 SSASY-INFEQPONPDNYLTPPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDAS 119
 Db 1026 YSEFGLGGIGVLPEDNYLITPALDLPGGKLTFWVCAQDANYASEHYAVYASSTGNDAS 1085

Qy 120 NFANALLEEVLTAKTVVTAPRAIGTRVQGTWYOKTVLPAGTKYVAFRHFQCTDFFWIN 179
 Db 1086 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYID 1143

Qy 180 LDEVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQDLWLTAHGGT 239
 Db 1144 LDEVEIKANGKRAADFTTFESSHGEAPAEWTTIDADGGQGLCLSSGQDLWLTAHGGT 1203

Qy 240 NVVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGPFQGDHYAVMISKTGTNAGDFT 299
 Db 1204 NVVASFWSNGMALNPNDNYLISKDVTGATKVKYYAVNDGPFQGDHYAVMISKTGTNAGDFT 1263

Qy 300 VVFETPNKNGKARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYIL 359
 Db 1264 VVFETPNKNGKARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYIL 1323

Qy 360 LDDIQFTMGSGPTPTDYTYTVYRDGTKIKEGLTTFEEDGVATGNHHEYCVVEKYTAGVS 419
 Db 1324 LDDIQFTMGSGPTPTDYTYTVYRDGTKIKEGLTTFEEDGVATGNHHEYCVVEKYTAGVS 1383

Qy 420 PKVCNVNTINPTQFNPNQNL 439
 Db 1384 PKVCNVNTINPTQFNPNQNL 1403

RESULT 3

ID P96967 PRELIMINARY; PRT; 1358 AA.
 AC P96967;
 DT 01-MAY-1997 (TREMELrel. 03, Created)
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hemagglutinin.
 GN HADG
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RA Han N., Lepine G., Whitlock J., Wojciechowski L., Progulski-Fox A.;
 RT "Cloning, sequencing and characterization of hagd, a member of the
 RT Harep multigene family in Porphyromonas gingivalis.";
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U68468; AAB49691.1; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.

DR SMART; SM00060; FN3; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1358 AA; 147102 MW; 47FCA0B25B06DED8 CRC64;
 Query Match 86.3%; Score 2060; DB 2; Length 1358;
 Best Local Similarity 88.6%; Pred. No. 1.8e-128;
 Matches 390; Conservative 12; Mismatches 32; Indels 6; Gaps 4;
 QY 1 GTPNPNPNPGTTTLSEFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICV 60
 DB 604 GTNPNPNPNPGTTTLSEFENGIPASWKTIDADGNGKPGNAPG---IAGNSNGCV 660
 QY 61 SSASY-INFGPQPNPNVLTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDAS 119
 DB 661 YSSEFGLGGIGLVLTDPNVLITPALDLANGKGLTFWVCAQDANYASEHYAVYASSTGNDAS 720
 QY 120 NFANALLEEVLTAKTVTAPEAIRGTVOGTYWQVTKVLPAGTKYVAFRHFQCTDFWIN 179
 DB 721 NFTNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYID 778
 QY 180 LDEVEIKANGKRADETFETFESSTHGEAPAEWTTIDADGGQGLCLSSGQLDMLTAHGGT 239
 DB 779 LDEVEIKANGKRADETFETFESSTHGEAPAEWTTIDADGGQGLDMLTAHGGT 838
 QY 240 NVVASFNGMALPNPNYLISKVGTGATKVKYKYAVNDGPGCDHYAVMISKTGTNAGDFT 299
 DB 839 NVVASFNGMALPNPNYLISKVGTGATKVKYKYAVNDGPGCDHYAVMISKTGTNAGDFT 898
 QY 300 VPEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHNCSDLNVL 359
 DB 899 VPEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHNCSDLNVL 958
 QY 360 LDDIQFTMGSSPTDITYTVYRDGTKIKEGLTETFEEDGVATGNHEVCVEVKYTAGVS 419
 DB 959 LDDIQFTMGSSPTDITYTVYRDGTKIKEGLTETFEEDGVATGNHEVCVEVKYTAGVS 1018
 QY 420 PKVCNVNTINPTQNPVNQL 439
 DB 1019 PKVCNVNTINPTQNPVNQL 1038
 RESULT 4
 Q9R9B7 PRELIMINARY; PRT; 1687 AA.
 ID Q9R9B7;
 AC Q9R9B7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hemagglutinin/protease.
 GN HAGE.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OC NCB1_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RA Han N., Dong H., Progluske-Fox A.;
 RT "Cloning and characterization of hage from P. gingivalis 381.";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF026946; AAD01810.1; -.
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; F:DNA recombination; IEA.
 DR GO; GO:0006281; F:DNA repair; IEA.
 DR GO; GO:0006260; F:DNA replication; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.

DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25_1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease.
 SQ SEQUENCE 1687 AA; 183702 MW; D085B516A399FE70 CRC64;
 Query Match 86.2%; Score 2058; DB 2; Length 1687;
 Best Local Similarity 88.8%; Pred. No. 3.2e-128;
 Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;
 QY 3 PNPENPNPGTTTLSEFENGIPASWKTIDADGNNWTTTTPPGGTSFAGHNSAICV 62
 DB 935 PNPENPNPGTTTLSEFENGIPASWKTIDADGNGKPGNAPG---IAGNSNGCVYS 991
 QY 63 ASY-INFGPQPNPNVLTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
 DB 992 ESFGLGGIGLVLTDPNVLITPALDLPNGKGLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051
 QY 122 ANALLEEVLTAKTVTAPEAIRGTVOGTYWQVTKVLPAGTKYVAFRHFQCTDFWINLD 181
 DB 1052 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQSTDMFYIDL 1109
 QY 182 EVEIKANGKRADETFETFESSTHGEAPAEWTTIDADGGQGLCLSSGQLDMLTAHGGT 241
 DB 1110 EVEIKANGKRADETFETFESSTHGEAPAEWTTIDADGGQGLCLSSGQLDMLTAHGGT 1169
 QY 242 VAFSWMNGMALPNPNYLISKVGTGATKVKYKYAVNDGPGCDHYAVMISKTGTNAGDFTV 301
 DB 1170 VAFSWMNGMALPNPNYLISKVGTGATKVKYKYAVNDGPGCDHYAVMISKTGTNAGDFTV 1229
 QY 302 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHNCSDLNVL 361
 DB 1230 FEETPNKNGGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHNCSDLNVL 1289
 QY 362 DIQFTMGSSPTDITYTVYRDGTKIKEGLTETFEEDGVATGNHEVCVEVKYTAGVSPK 421
 DB 1290 DIQFTMGSSPTDITYTVYRDGTKIKEGLTETFEEDGVATGNHEVCVEVKYTAGVSPK 1349
 QY 422 VCVNVNTINPTQNPVNQL 439
 DB 1350 ECVNVNTINPTQNPVNQL 1367
 RESULT 5
 Q51816 PRELIMINARY; PRT; 1704 AA.
 ID Q51816
 AC Q51816;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arg-gingipain-1 proteinase.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OC NCB1_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95138080; PubMed=7836351;
 RA Pavloff N., Potempa J., Pike R.N., Prochazka V., Kiefer M.C.,
 RA Travis J., Barr P.J.;
 RT "Molecular cloning and structural characterization of the Arg-gingipain proteinase of Porphyromonas gingivalis. Biosynthesis as a proteinase-adhesin polyprotein.";
 RL J. Biol. Chem. 270:1007-1010(1995).
 DR EMBL; U15282; AAA69539.1; -.
 DR PIR; A55426; A55426.
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; F:DNA recombination; IEA.
 DR GO; GO:0006281; F:DNA repair; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR001769; Peptidase_C25.

DR GO:0006281; P:DNA repair; IEA.
 DR GO:0006260; P:DNA replication; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000977; DNA ligase.
 DR InterPro: IPR007110; Ig-like.
 DR InterPro: IPR001769; Peptidase_C25.
 DR InterPro: IPR005536; Peptidase_C25_C.
 DR Pfam: PF01364; Peptidase_C25; 1.
 DR Pfam: PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 FT CHAIN 228 719
 MATURE 50-KDA CYSTEINE PROTEINASE
 GINGIPAIN.
 QY SEQUENCE 1704 AA; 185436 MW; 6A34B40131C2A676 CRC64;
 Best Local Similarity 86.2%; Score 2058; DB 2; Length 1704;
 Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;
 QY 3 PNPENPNPGTTTISEFENGIPASWKTIIDADGNNWTTTPPGGTSFAGHNSAICVSS 62
 Db PNPENPNPGTTTISEFENGIPASWKTIIDADGNGKPKGNAPG---IAGYNSNGCVYS 1008
 QY 63 ASY-INTEGPNPNYLVTPSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
 Db 1009 ESFGLGIGVLTDPNYLITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGNDASNF 1068
 QY 122 ANALLEVLTAKTVVTAPAIRTRVOGTWYKTVQVLPAGTKYVAFRHFPGCTDFFWNL 181
 Db 1069 TNALLEITITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYILD 1126
 QY 182 EVEIKANGKRAFFTETTESSTHGEAPAEWTTIDADGGQGWCLCSSGQLDMLTAHGGT 241
 Db 1127 EVEIKANGKRAFFTETTESSTHGEAPAEWTTIDADGGQGWCLCSSGQLDMLTAHGGT 1186
 QY 242 VASFSWNGMALNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVV 301
 Db 1187 VASFSWNGMALNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDFTVV 1246
 QY 302 FEETPNKINGGARFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLYILL 361
 Db 1247 FEETPNKINGGARFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLYILL 1306
 QY 362 DIQFTMGSPPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPK 421
 Db 1307 DIQFTMGSPPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGVSPK 1366
 QY 422 VCVNVTINPTQFNPVQNL 439
 Db 1367 ECVNVTINPTQFNPVQNL 1384

RESULT 6
 Q9ZNB5 PRELIMINARY; PRT; 1223 AA.
 ID Q9ZNB5
 AC Q9ZNB5;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 130K-HMGD (Fragment).
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 CX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=381;
 RX MEDLINE=99143166; PubMed=9988746;
 RA Shibata Y., Hayakawa M., Takiguchi H., Shiroza T., Abiko Y.;
 "Determination and characterization of the hemagglutinin-associated
 RT short motifs found in Porphyromonas gingivalis multiple gene
 RT products.";
 RL J. Biol. Chem. 274:5012-5020 (1999).
 DR EMBL; AB019363; BAA34341.1; -.

DR GO:0005524; F:ATP binding; IEA.
 DR GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO:0006310; P:DNA recombination; IEA.
 DR GO:0006281; P:DNA repair; IEA.
 DR GO:0006260; P:DNA replication; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000977; DNA ligase.
 DR InterPro: IPR001769; Peptidase_C25.
 DR InterPro: IPR005536; Peptidase_C25_C.
 DR Pfam: PF01364; Peptidase_C25; 1.
 DR Pfam: PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 FT NON TER 1
 QY SEQUENCE 1223 AA; 131542 MW; 00225CD2BA9F91B3 CRC64;
 Query Match 86.1%; Score 2056.5; DB 2; Length 1223;
 Best Local Similarity 88.7%; Pred. No. 2.7e-128;
 Matches 391; Conservative 11; Mismatches 32; Indels 7; Gaps 5;
 QY 1 GTPNPNPNPGT-TTISEFENGIPASWKTIIDADGNNWTTTPPGGTSFAGHNSAIC 59
 Db 460 GTNPNPNPNPGT-TTISEFENGIPASWKTIIDADGNGKPKGNAPG---IAGYNSNGC 516
 QY 60 VSSASY-INTEGPNPNYLVTPSLPGGGTLTFWVCAQDANYASEHYAVYASSTGND 118
 Db 517 VYSEFGLGIGVLTDPNYLITPALDLPNGGKLTFFWVCAQDANYASEHYAVYASSTGND 576
 QY 119 SNFANALLEVLTAKTVVTAPAIRTRVOGTWYKTVQVLPAGTKYVAFRHFPGCTDFFWI 178
 Db 577 SNTNALLEITITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYI 634
 QY 179 NLDEVEIKANGKRAFFTETTESSTHGEAPAEWTTIDADGGQGWCLCSSGQLDMLTAHGG 238
 Db 635 DLDEVEIKANGKRAFFTETTESSTHGEAPAEWTTIDADGGQGWCLCSSGQLDMLTAHGG 694
 QY 239 TNVVASFSWNGMALNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 298
 Db 695 TNVVASFSWNGMALNPNYLISKDVTGATKVKYVAVNDGPGDHYAVMISKTGTNAGDF 754
 QY 299 TVVFEETPNKINGGARFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLYNI 358
 Db 755 TVVFEETPNKINGGARFGLSTEADGAKPOSVMERTVDLPAGTKYVAFRHYNCSDLYNI 814
 QY 359 LLDDIQFTMGSPPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGV 418
 Db 815 LLDDIQFTMGSPPTDYYTVYRDGTKIKEGLTETTFEEDGVATGNHEVCVEVKYTAGV 874
 QY 419 SPKVCVNVVTINPTQFNPVQNL 439
 Db 875 SPKVCVNVVTINPTQFNPVQNL 895

RESULT 7
 Q51839 PRELIMINARY; PRT; 1706 AA.
 ID Q51839
 AC Q51839; Q51840;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Arginine-specific thiol protease precursor.
 GN PRTR.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 CX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RX MEDLINE=95160709; PubMed=7857299;
 RA Kirsbaum L., Sotiropoulos C., Jackson C., Cleal S., Slakeski N.,
 RA Reynolds E.C.;
 RT "Complete nucleotide sequence of a gene prTR of Porphyromonas

RT gingivalis W50 encoding a 132 kDa protein that contains an arginine-specific thiol endopeptidase domain and a haemagglutinin domain.";
 RL Biochem. Biophys. Res. Commun. 207:424-431(1995).
 RN [2]
 RC
 RP SEQUENCE FROM N.A.
 RX STRAIN=W50;
 RX MEDLINE=96311339; PubMed=8713096;
 RA Slakeski N., Cleal S.M., Reynolds E.C.;

RT "Characterization of a Porphyromonas gingivalis gene prtr that encodes an arginine-specific thiol proteinase and multiple adhesins.";
 RL Biochem. Biophys. Res. Commun. 224:605-610(1996).
 RN [3]
 RC
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Reynolds E.;

RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RC
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RA Slakeski N.;

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L26341; AAC18876.1; -.
 DR HSSP; P95493; 1CVR.
 DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.

DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; 1.
 DR Pfam; PF03785; Peptidase_C25_C; 1.

DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Protease; Signal.
 FT SIGNAL
 FT CHAIN 1 23 POTENTIAL.
 FT CHAIN 228 1706 ARGININE-SPECIFIC THIOI PROTEASE.

SQ SEQUENCE 1706 AA; 185626 MW; E8BDF07C9813B844 CRC64;
 Query Match 85.8%; Score 2048; DB 2; Length 1706;
 Best Local Similarity 88.1%; Pred. No. 1.5e-127;
 Matches 386; Conservative 14; Mismatches 32; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTSESFENGIPASWKTIDADGNGNWTTPPPGCTSFAGHNSAICVSS 62
 DB 954 PNPENPNPGTTTSESFENGIPASWKTIDADGNGHGWKPGNAPG---IAGVNSNGCVYS 1010

QY 63 ASY-INFGPQPNPNYLVTPPELPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
 DB 1011 ESFGLGGIGLVLPDNYLTIPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

QY 122 ANALLEEVLTAKTVVTAPEAIGRTVQGTWQKTVQVLPAGTKYVAFRHCCTDFWINDL 181
 DB 1071 TNALLEETITAGK-VRSPEAMRG-RIQGTWRQKTVQVLPAGTKYVAFRHCCTDFWINDL 1128

QY 182 EYEIKANGKRAFTTETFEESTHGEAPAEWTTTIDADGGGWLCLSSGQDLWLTAGGTNV 241
 DB 1129 EYEIKANGKRAFTTETFEESTHGEAPAEWTTTIDADGGGWLCLSSGQDLWLTAGGTNV 1188

QY 242 VASFSWNGMALPNPNYLSKDVGTATKVKYKYAVNDGPFDDHYAVMISKTGNAGDTVV 301
 DB 1189 VSSFSWNGMALPNPNYLSKDVGTATKVKYKYAVNDGPFDDHYAVMISKTGNAGDTVV 1248

QY 302 FEETPNKNGKARGLSLEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLYILLD 361
 DB 1249 FEETPNKNGKARGLSLEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLYILLD 1308

QY 362 DIQFTMGSSPTTDYTYTYVRDGTIKKGLTETTFEEDGVAIGNHEYCEVVEKYTAGVSPK 421
 DB 1309 DIQFTMGSSPTTDYTYTYVRDGTIKKGLTETTFEEDGVAIGNHEYCEVVEKYTAGVSPK 1368

QY 422 VCVNVTINPTQFNPNVQNL 439
 DB 1369 KCVNVTINPTQFNPNVQNL 1386

RESULT 8

Q51838 PRELIMINARY; PRT; 1706 AA.
 AC Q51838;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Protease precursor.
 GN PRP1.

OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (Class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OC NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=W50;
 RX MEDLINE=96071894; PubMed=7591131;
 RA Aduse-Opoku J., Muir J., Slaney J.M., Rangarajan M., Curtis M.A.;

RT "Characterization, genetic analysis, and expression of a protease antigen (Prp1) of Porphyromonas gingivalis W50.";
 RL Infect. Immun. 63:4744-4754(1995).
 RN [2]
 RC
 RP SEQUENCE FROM N.A.

RC STRAIN=W50;
 RA Rangarajan M., Aduse-Opoku J., Slaney J.M., Young K.A., Curtis M.A.;

RT "The prp1 and the pr2 arginine-specific protease genes of Porphyromonas gingivalis W50 produce five biochemically distinct enzymes.";
 RL Mol. Microbiol. 23:0-0(1997).

DR EMBL; X82680; CAA57997.1; -.
 DR HSSP; P95493; 1CVR.
 DR MEROPS; C25.001; -.

DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.

DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA_ligase.
 DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR001769; Peptidase_C25.
 DR InterPro; IPR005536; Peptidase_C25_C.
 DR Pfam; PF01364; Peptidase_C25; 1.

DR Pfam; PF03785; Peptidase_C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Signal.
 FT SIGNAL
 FT CHAIN 1 23 POTENTIAL.
 FT CHAIN 228 719 ALPHA-PROTEASE.

FT CHAIN 720 1262 BETA-ADHESIN.
 SQ SEQUENCE 1706 AA; 185705 MW; 0E56DCD87FDA8CDD CRC64;

Query Match 85.3%; Score 2037; DB 2; Length 1706;
 Best Local Similarity 87.9%; Pred. No. 8.2e-127;
 Matches 385; Conservative 13; Mismatches 34; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTSESFENGIPASWKTIDADGNGNWTTPPPGCTSFAGHNSAICVSS 62
 DB 954 PNPENPNPGTTTSESFENGIPASWKTIDADGNGHGWKPGNAPG---IAGVNSNGCVYS 1010

QY 63 ASY-INFGPQPNPNYLVTPPELPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
 DB 1011 ESFGLGGIGLVLPDNYLTIPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1070

QY 122 ANALLEEVLTAKTVVTAPEAIGRTVQGTWQKTVQVLPAGTKYVAFRHCCTDFWINDL 181
 DB 1071 TNALLEETITAGK-VRSPEAIRG-RIOSTWRQKTVQVLPAGTKYVAFRHCCTDFWINDL 1128

QY 182 EVEIKANGKADFTETTESGTHGEAPAEWTTIDADGGQGWMLCLSSGQDMLTAHGGTNN 241
 DB 1129 EVEIKANGKADFTETTESGTHGEAPAEWTTIDADGGQGWMLCLSSGQDMLTAHGGTNN 1188
 QY 242 VASFSMNGMALPNPNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKTGTNAGDFTVV 301
 DB 1189 VSSFSMNGMALPNPNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKTGTNAGDFTVV 1248
 QY 302 FEETPGINGKKGARFGLSTADGAKPQSWIERTVLDLPAGTKVYAFPHYNCSDLNILYLD 361
 DB 1249 FEETPGINGKKGARFGLSTADGAKPQSWIERTVLDLPAGTKVYAFPHYNCSDLNILYLD 1308
 QY 362 DIQFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCDEVKVTAGVSPK 421
 DB 1309 DIQFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCDEVKVTAGVSPK 1368
 QY 422 VCNVNTINPTQFNPQNL 439
 DB 1369 KCNVNTVNSTQFNPQNL 1386

RESULT 9
 ID P72196 PRELIMINARY; PRT; 1097 AA.
 AC P72196;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE TonB-linked adhesin precursor.
 GN TLA.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97386416; PubMed=9244265;
 RA Aduse-Opoku J., Slaney J.M., Young K.A., Muir J., Rangarajan M.,
 Curtis M.A.;
 RT "The tla gene of Porphyromonas gingivalis W50: a homologue of the
 RT arginine-specific protease precursor (PrpR) which shares sequence
 RT similarity to TonB-linked receptors.";
 RL J. Bacteriol. 179:4778-4788 (1997).
 CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; Y07618; CAA68897.1; -;
 DR GO; GO:0019867; C:outer membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR Pfam; PF00593; TonB dep Rec; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 KW Membrane; Outer membrane; Receptor; Signal; TonB box.
 FT SIGNAL 1 53 POTENTIAL.
 SQ SEQUENCE 1097 AA; 118731 MW; 73BBA37B421F8B9 CRC64;

Query Match 85.0%; Score 2030.5; DB 2; Length 1097;
 Best Local Similarity 87.3%; Pred. No. 1.2e-126;
 Matches 385; Conservative 15; Mismatches 34; Indels 7; Gaps 5;

QY 1 GTPNPNPNPQGT-TTILSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAIC 59
 DB 334 GTPNPNPNPNPQGT-TTILSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAIC 390
 QY 60 VSSASY-INTEGPQPNPNLYLTPGLSGGTLTFWVCAQDANYASEHYAVASSTGND 118

DB 391 VYSEFGLGIGVLPDNYLIITPALDPLNGKILTFWVCAQDANYASEHYAVASSTGND 450
 QY 119 SNFANALLEEVLTAKTVTAPPAIRGTRVQGTWYOKTVOLPAGTKVYAFPHYNCSDLNILYLD 178
 DB 451 SNFTNALLBETITAKG-VRSPKAIIRG-RIOGTWROKTVOLPAGTKVYAFPHYNCSDLNILYLD 508
 QY 179 NLDEVEIKANGKADFTETTESGTHGEAPAEWTTIDADGGQGWMLCLSSGQDMLTAHGG 238
 DB 509 DLDEVEIKANGKADFTETTESGTHGEAPAEWTTIDADGGQGWMLCLSSGQDMLTAHGG 568
 QY 239 TNNVASFSMNGMALPNPNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKTGTNAGD 298
 DB 569 SNVVSFSMNGMALPNPNLYLSKDVGTGATKVKYKYAVNDGFGPDGHYAVMISKTGTNAGD 628
 QY 299 TVVFEETPGINGKKGARFGLSTADGAKPQSWIERTVLDLPAGTKVYAFPHYNCSDLNILYLD 358
 DB 629 TVVFEETPGINGKKGARFGLSTADGAKPQSWIERTVLDLPAGTKVYAFPHYNCSDLNILYLD 688
 QY 359 LLDDIQTFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCDEVKVTAGV 418
 DB 689 LLDDIQTFTMGSSPTDYYTVYRDGKIKEGLTETTFEEDGVATGNHGYCDEVKVTAGV 748
 QY 419 SPKVCNVNTINPTQFNPQNL 439
 DB 749 SPKKVCNVNTVNSTQFNPQNL 769

RESULT 10
 ID 007442 PRELIMINARY; PRT; 1732 AA.
 AC 007442;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Lysine-specific cysteine proteinase.
 GN PrkK.
 OS Porphyromonas gingivalis (Bacteroides gingivalis).
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Porphyromonadaceae; Porphyromonas.
 OX NCBI_TaxID=837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W50;
 RX MEDLINE=99235907; PubMed=10219167;
 RA Slakeski N., Cleal S.M., Bhogal P.S., Reynolds E.C.;
 RT "Characterization of a Porphyromonas gingivalis gene prtK that encodes
 RT a lysine-specific cysteine proteinase and three sequence-related
 RT adhesins.";
 RL Oral Microbiol. Immunol. 14:92-97 (1999).
 DR EMBL; U75366; AAB60809.1; -;
 DR MEROPS; C25.002; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008234; F:cysteine-type peptidase activity; IEA.
 DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006281; P:DNA repair; IEA.
 DR GO; GO:0006260; P:DNA replication; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR001769; Peptidase C25.
 DR InterPro; IPR005536; Peptidase C25_C.
 DR Pfam; PF01364; Peptidase C25; 1.
 DR Pfam; PF03785; Peptidase C25_C; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 SQ SEQUENCE 1732 AA; 187914 MW; 45D5B91377391703 CRC64;

Query Match 85.0%; Score 2030.5; DB 2; Length 1732;
 Best Local Similarity 87.3%; Pred. No. 2.3e-126;
 Matches 385; Conservative 15; Mismatches 34; Indels 7; Gaps 5;

QY 1 GTPNPNPNPQGT-TTILSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAIC 59
 DB 969 GTPNPNPNPNPQGT-TTILSESFENGIPASWKTIDADGGNNWTTTPPPGGTSFAGHNSAIC 1025


```
Best Local Similarity 86.6%; Pred. No. 4.2e-125;
Matches 382; Conservative 16; Mismatches 36; Indels 7; Gaps 5;

QY 1 GTPNPNPNPGT-TTLESFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAIC 59
Db 969 GTPNPNPNPNPOTLSEFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAIC 1025
QY 60 VSSASY-INFEQPNPNYLTPELSLPGGGTTLTFWVCAQDANYASHEHYAVYASSTGND 118
Db 1026 VYSEFGLGGVLTPTDNYLTALDLPNGKLTFFWVCAQDANYASHEHYAVYASSTGND 1085
QY 119 SNFANALLEVLAKTAVTAPAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFPGCTDFWI 178
Db 1086 SNFTNALLEETITAKG-VRSFKAIRG-RIQGTWRQKTVQVLPAGTKYVAFRHFPGCTDFWI 1143
QY 179 NLDEVEIKANGKRAFTETPESSTHGPAPAEWTTIDADGGQGWLCSSGQDLWLTAGG 238
Db 1144 DLDEVEIKANGKRAFTETPESSTHGPAPAEWTTIDADGGQGWLCSSGQDLWLTAGG 1203
QY 239 TNVVASFWSMGALNPYLSKDVTKATKVKYVAVNDGFPDGHYAVMISKTGTNAGDF 298
Db 1204 SNVVSFSWSMGALNPYLSKDVTKATKVKYVAVNDGFPDGHYAVMISKTGTNAGDF 1263
QY 299 TVVFEETPNGKNGARFGLSTEADGAKPOSVMWERTVQVLPAGTKYVAFRHFPGCTDFWI 358
Db 1264 TVVFEETPNGKNGARFGLSTEADGAKPOSVMWERTVQVLPAGTKYVAFRHFPGCTDFWI 1323
QY 359 LLDIIQFTMGSSPTPTDYTVVYRDGFKIKEGLTETTFEEDGVATGNHCEVVEKYTAGV 418
Db 1324 LLDIIQFTMGSSPTPTDYTVVYRDGFKIKEGLTETTFEEDGVATGNHCEVVEKYTAGV 1383
QY 419 SPKVCVNTINPTQFNPVQNL 439
Db 1384 SPKVCVNTINPTQFNPVQNL 1404

RESULT 13
Q9FAJ0 PRELIMINARY; PRT; 925 AA.
ID Q9FAJ0
AC Q9FAJ0
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative outer membrane protein P57.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Czajkowski L., Hocking D., Margetts M., Webb E., Rothel L.,
RA Patterson M., Agius C., Camuglia S., Reynolds E., Barr I.G.;
RT "Identification of vaccine candidates from genomic analysis of
RT Porphyromonas gingivalis";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007522; AAG24228.1; -
DR InterPro; IPR003961; FN_III.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 925 AA; 103632 MW; 5FF2198D6914DAE8 CRC64;

Query Match 29.0%; Score 691.5; DB 2; Length 925;
Best Local Similarity 30.0%; Pred. No. 1.3e-37;
Matches 175; Conservative 63; Mismatches 138; Indels 207; Gaps 20;

QY 16 LSESFENGIPASWKTIADGDNWTTTPPGG-----TSPFAGHNSAICVSSASINF 68

Best Local Similarity 86.6%; Pred. No. 4.2e-125;
Matches 382; Conservative 16; Mismatches 36; Indels 7; Gaps 5;

QY 1 GTPNPNPNPGT-TTLESFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAIC 59
Db 969 GTPNPNPNPNPOTLSEFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAIC 1025
QY 60 VSSASY-INFEQPNPNYLTPELSLPGGGTTLTFWVCAQDANYASHEHYAVYASSTGND 118
Db 1026 VYSEFGLGGVLTPTDNYLTALDLPNGKLTFFWVCAQDANYASHEHYAVYASSTGND 1085
QY 119 SNFANALLEVLAKTAVTAPAIRGRVQGTWYQKTVQVLPAGTKYVAFRHFPGCTDFWI 178
Db 1086 SNFTNALLEETITAKG-VRSFKAIRG-RIQGTWRQKTVQVLPAGTKYVAFRHFPGCTDFWI 1143
QY 179 NLDEVEIKANGKRAFTETPESSTHGPAPAEWTTIDADGGQGWLCSSGQDLWLTAGG 238
Db 1144 DLDEVEIKANGKRAFTETPESSTHGPAPAEWTTIDADGGQGWLCSSGQDLWLTAGG 1203
QY 239 TNVVASFWSMGALNPYLSKDVTKATKVKYVAVNDGFPDGHYAVMISKTGTNAGDF 298
Db 1204 SNVVSFSWSMGALNPYLSKDVTKATKVKYVAVNDGFPDGHYAVMISKTGTNAGDF 1263
QY 299 TVVFEETPNGKNGARFGLSTEADGAKPOSVMWERTVQVLPAGTKYVAFRHFPGCTDFWI 358
Db 1264 TVVFEETPNGKNGARFGLSTEADGAKPOSVMWERTVQVLPAGTKYVAFRHFPGCTDFWI 1323
QY 359 LLDIIQFTMGSSPTPTDYTVVYRDGFKIKEGLTETTFEEDGVATGNHCEVVEKYTAGV 418
Db 1324 LLDIIQFTMGSSPTPTDYTVVYRDGFKIKEGLTETTFEEDGVATGNHCEVVEKYTAGV 1383
QY 419 SPKVCVNTINPTQFNPVQNL 439
Db 1384 SPKVCVNTINPTQFNPVQNL 1404

RESULT 14
Q9KIB3 PRELIMINARY; PRT; 312 AA.
ID Q9KIB3
AC Q9KIB3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical outer membrane protein PG27.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237555; AAF81413.1; -
SQ SEQUENCE 312 AA; 34592 MW; 0D5792C9643A25F5 CRC64;

Query Match 23.9%; Score 571.5; DB 2; Length 312;
Best Local Similarity 33.3%; Pred. No. 3e-30;
Matches 138; Conservative 23; Mismatches 71; Indels 183; Gaps 8;

QY 16 LSESFENGIPASWKTIADGDNWTTTPPGGTSFAGHNSAICVSSASINFEGQNPD 75
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:38:40 ; Search time 13.5635 Seconds
(without alignments)
1670.936 Million cell updates/sec

Title: US-08-570-311-22

Perfect score: 2388

Sequence: 1 GPNPNPNPNPCTTTLSSEF.....PKVCNVNTINPTQFNPVQNL 439

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pap.*
- 2: /cgn2_6/prodata/2/iaa/5A COMB.pap.*
- 3: /cgn2_6/prodata/2/iaa/5A COMB.pap.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pap.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pap.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 2388 | 100.0 | 439 | 2 | US-08-570-311-22 |
| 2 | 2388 | 100.0 | 2628 | 2 | US-08-570-311-14 |
| 3 | 2381 | 98.5 | 456 | 2 | US-08-570-311-18 |
| 4 | 2351 | 98.5 | 456 | 2 | US-08-570-311-20 |
| 5 | 2321 | 97.2 | 450 | 2 | US-08-570-311-16 |
| 6 | 2058 | 86.2 | 1477 | 4 | US-09-482-500A-1 |
| 7 | 2058 | 86.2 | 1687 | 4 | US-08-570-311-29 |
| 8 | 2058 | 86.2 | 1704 | 3 | US-08-336-308A-10 |
| 9 | 2058 | 86.2 | 1704 | 3 | US-08-822-324-6 |
| 10 | 2058 | 86.2 | 1704 | 3 | US-09-490-931-10 |
| 11 | 2055 | 86.1 | 1087 | 2 | US-08-570-311-8 |
| 12 | 2055 | 86.1 | 1087 | 2 | US-08-353-485-8 |
| 13 | 2055 | 86.1 | 1358 | 2 | US-08-570-311-27 |
| 14 | 2048 | 85.8 | 1706 | 4 | US-09-066-330-10 |
| 15 | 2030.5 | 85.0 | 1732 | 4 | US-09-066-330-11 |
| 16 | 2025.5 | 84.8 | 1732 | 2 | US-08-570-311-10 |
| 17 | 2025.5 | 84.8 | 1732 | 2 | US-08-353-485-10 |
| 18 | 458.5 | 19.2 | 497 | 2 | US-08-570-311-2 |
| 19 | 458.5 | 19.2 | 497 | 2 | US-08-353-485-2 |
| 20 | 294.5 | 12.3 | 942 | 1 | US-08-141-324-14 |
| 21 | 294.5 | 12.3 | 942 | 1 | US-08-541-902-14 |
| 22 | 242 | 10.1 | 49 | 3 | US-08-822-324-18 |
| 23 | 195 | 8.2 | 46 | 3 | US-08-822-324-9 |
| 24 | 131.5 | 5.5 | 1638 | 4 | US-09-071-035-258 |
| 25 | 131.5 | 5.5 | 1638 | 4 | US-09-071-035-262 |
| 26 | 131.5 | 5.5 | 1638 | 4 | US-09-071-035-266 |
| 27 | 131.5 | 5.5 | 1747 | 4 | US-09-134-000C-5999 |

28 129 5.4 24 4 US-09-066-330-6 Sequence 6, Appli
29 128 5.4 25 3 US-08-822-324-19 Sequence 19, Appli
30 125.5 5.3 669 4 US-09-071-035-264 Sequence 264, App
31 125 5.2 25 3 US-08-822-324-14 Sequence 14, Appli
32 124 5.2 2315 4 US-09-543-681A-5434 Sequence 5434, Ap
33 123.5 5.2 1095 4 US-08-851-567B-34 Sequence 34, Appli
34 120 5.0 1833 4 US-08-621-944A-4 Sequence 4, Appli
35 120 5.0 1833 4 US-08-945-567D-4 Sequence 4, Appli
36 120 5.0 1992 4 US-08-621-944A-3 Sequence 3, Appli
37 120 5.0 1992 4 US-08-945-567D-3 Sequence 3, Appli
38 118.5 5.0 2123 3 US-08-968-685A-10 Sequence 10, Appli
39 118 4.9 24 3 US-08-822-324-20 Sequence 20, Appli
40 118 4.9 2353 3 US-09-377-155-33 Sequence 33, Appli
41 118 4.9 2353 3 US-08-913-942-4 Sequence 4, Appli
42 118 4.9 2353 4 US-09-669-974-33 Sequence 33, Appli
43 118 4.9 2353 4 US-09-797-862-33 Sequence 33, Appli
44 118 4.9 2411 4 US-09-268-347-36 Sequence 36, Appli
45 117.5 4.9 331 1 US-08-535-230A-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-570-311-22
; Sequence 22, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:

APPLICANT: Progulsk-Fox, Ann
APPLICANT: Tunwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESSES:

ADDRESSES: Ted W. Whitlock

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL USA

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/570,311

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/353,485

FILING DATE: 09-DEC-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/647,119

FILING DATE: 25-JAN-1991

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/241,640

FILING DATE: 08-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.

REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: UF15.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 439 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-22

Query Match      100.0%; Score 2388; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.7e-204;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPENPNPGTTTLLSEFENGIPASWKTIIDAGDGNWTTTTPPGGTSFAGHNSAICV 60
Db 1 GTPNPENPNPGTTTLLSEFENGIPASWKTIIDAGDGNWTTTTPPGGTSFAGHNSAICV 60

QY 61 SSASYINFEQPNPDNYLVTPELSPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 61 SSASYINFEQPNPDNYLVTPELSPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120

QY 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWINL 180
Db 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWINL 180

QY 181 DEVEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQWCLSSGQDMLTAHGGTN 240
Db 181 DEVEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQWCLSSGQDMLTAHGGTN 240

QY 241 VVASFWSNGMALNPNDNYLISKDVTKATKVKYKYAVVNDGFGPDGHYAVNMISKTGTNAGDFTV 300
Db 241 VVASFWSNGMALNPNDNYLISKDVTKATKVKYKYAVVNDGFGPDGHYAVNMISKTGTNAGDFTV 300

QY 301 VFETPNKNGKARFGLSTEADGAKQSVWIERTVDLPAKTGYVAFRHHGCTDFFWINL 360
Db 301 VFETPNKNGKARFGLSTEADGAKQSVWIERTVDLPAKTGYVAFRHHGCTDFFWINL 360

QY 361 DDIOFTMGSGSPFDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVVEVKYTAGVSP 420
Db 361 DDIOFTMGSGSPFDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVVEVKYTAGVSP 420

QY 421 KVCNVVTINPTQFNPVQNL 439
Db 421 KVCNVVTINPTQFNPVQNL 439

RESULT 2
US-08-570-311-14
; Sequence 14, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-570-311-14

Query Match      100.0%; Score 2388; DB 2; Length 2628;
Best Local Similarity 100.0%; Pred. No. 4.2e-203;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTPNPENPNPGTTTLLSEFENGIPASWKTIIDAGDGNWTTTTPPGGTSFAGHNSAICV 60
Db 1862 GTNPENPNPGTTTLLSEFENGIPASWKTIIDAGDGNWTTTTPPGGTSFAGHNSAICV 1921

QY 61 SSASYINFEQPNPDNYLVTPELSPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 120
Db 1922 SSASYINFEQPNPDNYLVTPELSPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASN 1981

QY 121 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWINL 180
Db 1982 FANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQVLPAGTKYVAFRHHGCTDFFWINL 2041

QY 181 DEVEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQWCLSSGQDMLTAHGGTN 240
Db 2042 DEVEIKANGKRAADTFETFEFSTHGEAPAEWTTIDADGGQWCLSSGQDMLTAHGGTN 2101

QY 241 VVASFWSNGMALNPNDNYLISKDVTKATKVKYKYAVVNDGFGPDGHYAVNMISKTGTNAGDFTV 300
Db 2102 VVASFWSNGMALNPNDNYLISKDVTKATKVKYKYAVVNDGFGPDGHYAVNMISKTGTNAGDFTV 2161

QY 301 VFETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAKTGYVAFRHHGCTDFFWINL 360
Db 2162 VFETPNKNGKARFGLSTEADGAKPQSVWIERTVDLPAKTGYVAFRHHGCTDFFWINL 2221

QY 361 DDIOFTMGSGSPFDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVVEVKYTAGVSP 420
Db 2222 DDIOFTMGSGSPFDYTYTVYRDGTKIKEGLTETFEEDGVATGNHEYCVVEVKYTAGVSP 2281

QY 421 KVCNVVTINPTQFNPVQNL 439
Db 2282 KVCNVVTINPTQFNPVQNL 2300

RESULT 3
US-08-570-311-18
; Sequence 18, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
```

APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSES: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-18

Query Match 98.5%; Score 2351; DB 2; Length 456;
Best Local Similarity 98.2%; Pred. No. 5.7e-201;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTPNPNPNGPCTTTLSEFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
DB 1 GTPNPNPNGPCTTTLSEFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

QY 61 SSASINFEQPNPNLYLTPLESLPFGGTLTFWCAQDANYASEHYAVYASSTGNDASN 120
DB 61 SSASINFEQPNPNLYLTPLESLPFGGTLTFWCAQDANYASEHYAVYASSTGNDASN 120

QY 121 FANALLEVLTAKTVTVAPEATRGTRVQGTWQKTVQKTVQKTVQKTVQKTVQKTVQKTVQKTV 180
DB 121 FANALLEVLTAKTVTVAPEATRGTRVQGTWQKTVQKTVQKTVQKTVQKTVQKTVQKTVQKTV 180

QY 181 DEVEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTN 240
DB 181 DVEIKANGKRAADFTETTESSTHGEAPAEWTTIDADGGQGWCLSSGQLDMLTAHGGTN 240

QY 241 VVASFSWNGMALPNPNYLISKDVTGATKVKYVAVNDGPPGHYAVMISKTGTNAGDFTV 300
DB 241 VVASFSWNGMALPNPNYLISKDVTGATKVKYVAVNDGPPGHYAVMISKTGTNAGDFTV 300

QY 301 VFEETPNGINKGARGFLSTEADGAKPOSVWERTVDLPAGTKYVAFRHYNCSDLNYILL 360
DB 301 VFEETPNGINKGARGFLSTEADGAKPOSVWERTVDLPAGTKYVAFRHYNCSDLNYILL 360

QY 361 DDIOFTMGSPPTDYYTVYVRDGTGKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSP 420
DB 361 DDIOFTMGSPPTDYYTVYVRDGTGKIKEGLTETTFEEDGVATGNHEYCVKVTAGVSP 420

QY 421 KCVNVNTINPTQFNPQNL 439
DB 421 KECNVTVDPVQFNPQNL 439

RESULT 4
US-08-570-311-20
Sequence 20, Application US/08570311
Patent No. 5824791
GENERAL INFORMATION:
APPLICANT: Progulske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-20

Query Match 98.5%; Score 2351; DB 2; Length 456;
Best Local Similarity 98.2%; Pred. No. 5.7e-201;
Matches 431; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTPNPNPNGPCTTTLSEFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICV 60
DB 1 GTPNPNPNGPCTTTLSEFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICA 60

```

QY 61 SSASYINFEGQPNPDNVLTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASN 120
DB 61 SSASYINFEGQPNPDNVLTPELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASN 120
QY 121 FANALLEVLTAKTVVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
DB 121 FANALLEVLTAKTVVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNL 180
QY 181 DEVEIKANGRADFTETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDLWLTAGGTTN 240
DB 181 DVEIKANGRADFTETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDLWLTAGGTTN 240
QY 241 VVASFSMNGMALPNPNLISKDVTKYKYVAVNDGFPQGDHYAVMISKTGNTAGDFTV 300
DB 241 VVASFSMNGMALPNPNLISKDVTKYKYVAVNDGFPQGDHYAVMISKTGNTAGDFTV 300
QY 301 VFETPNKNGGARFGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNVL 360
DB 301 VFETPNKNGGARFGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNVL 360
QY 361 DDIOFTMGSGPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
DB 361 DDIOFTMGSGPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSP 420
QY 421 KVCNVNTPNQFNPQNL 439
DB 421 KECNVNTPDQFNPQNL 439

RESULT 5
US-08-570-311-16
; Sequence 16, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Proguiske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.

```

```

; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UFL5.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-16

Query Match 97.2%; Score 2321; DB 2; Length 450;
Best Local Similarity 98.4%; Pred. No. 2.6e-198;
Matches 426; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 PNPNGTTLSSEFENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAICVSSASYI 66
DB 1 PNPNGTTLSSEFENGIPASWKTIDADGDNWNTTTPPGGTSFAGHNSAICVSSASYI 60
QY 67 NFESGPONPNLYLTPBELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASNANALL 126
DB 61 NFESGPONPNLYLTPBELSLPGGGLTFWVCAQDANYASEHYAVYASSTGNDASNANALL 120
QY 127 EEVLTAKTVVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNLDEVEIK 186
DB 121 EEVLTAKTVVTVAPEAIRGTRVQGTWYQKTVLPAGTKYVAFRHFCTDFFWNLDEVEIK 180
QY 187 ANGKRADFTETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDLWLTAGGTTNVSF 246
DB 181 ANGKRADFTETFEFSTHGEAPAEWTTIDADGGQGWLCSSGQDLWLTAGGTTNVSF 240
QY 247 WNGMALPNPNLISKDVTKYKYVAVNDGFPQGDHYAVMISKTGNTAGDFTVFEETP 306
DB 241 WNGMALPNPNLISKDVTKYKYVAVNDGFPQGDHYAVMISKTGNTAGDFTVFEETP 300
QY 307 NGKKGAREGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNVLDDIQFT 366
DB 301 NGKKGAREGLSTEADGAKPQSVMIERTVLDLPAGTKYVAFRHYNCSDLNVLDDIQFT 360
QY 367 MGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKVCNV 426
DB 361 MGSSPTPDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPKVCNV 420
QY 427 TINPTQFNPQNL 439
DB 421 TVDPVQFNPQNL 433

RESULT 6
US-09-482-500A-1
; Sequence 1, Application US/09482500A
; Patent No. 6627193
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Imamura, Takahisa
; APPLICANT: Potempa, Jan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROL OF BLOOD COAGULATION
; FILE REFERENCE: 235.00160101
; CURRENT APPLICATION NUMBER: US/09/482,500A
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/115,869
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-09-482-500A-1

Query Match 86.2%; Score 2058; DB 4; Length 1477;

```

```
Best Local Similarity 88.8%; Pred. No. 4.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKIIDADGDNWNTTTPPGGTSPAGHNSAICVSS 62
DB 725 PNPENPNPGTTLTSESFENGIPASWKIIDADGDNWNTTTPPGGTSPAGHNSAICVSS 781
QY 63 ASY-INFGPQPNPNLYLTPPELSLPGGGTITFWVCAQDANYASEHYAVYASSTGNDASNF 121
DB 782 ESFGLGGIGVLTDPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 841
QY 122 ANALLEEVLTAKTVTAPEAIRGTRVQGTWYKTVQVLPAGTKYVAFRHYNCSDLYILLD 181
DB 842 TNALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHYNCSDLYILLD 899
QY 182 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDMLTAHGGTNNV 241
DB 900 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDMLTAHGGTNNV 959
QY 242 VASFSWNGMALPNPNLYLSKDVGTGATKVKYVAVNDGFPGDHYAVMI SKGTGNAGDFTTV 301
DB 960 VASFSWNGMALPNPNLYLSKDVGTGATKVKYVAVNDGFPGDHYAVMI SKGTGNAGDFTTV 1019
QY 302 FEETPENGKNGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYILLD 361
DB 1020 FEETPENGKNGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYILLD 1079
QY 362 DIQFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
DB 1080 DIQFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1139
QY 422 VCVNVTINPTQFNPNVNL 439
DB 1140 ECVNVTINPTQFNPNVNL 1157

RESULT 7
US-08-570-311-29
; Sequence 29, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progukske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF INVENTIONS: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-570-311-29

Query Match 86.2%; Score 2058; DB 2; Length 1687;
Best Local Similarity 88.8%; Pred. No. 5.2e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTLTSESFENGIPASWKIIDADGDNWNTTTPPGGTSPAGHNSAICVSS 62
DB 935 PNPENPNPGTTLTSESFENGIPASWKIIDADGDNWNTTTPPGGTSPAGHNSAICVSS 991
QY 63 ASY-INFGPQPNPNLYLTPPELSLPGGGTITFWVCAQDANYASEHYAVYASSTGNDASNF 121
DB 992 ESFGLGGIGVLTDPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASSTGNDASNF 1051
QY 122 ANALLEEVLTAKTVTAPEAIRGTRVQGTWYKTVQVLPAGTKYVAFRHYNCSDLYILLD 181
DB 1052 TNALLEETITAGK-VRSPEAIRG-RIQGTWRQKTVDLPAGTKYVAFRHYNCSDLYILLD 1109
QY 182 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDMLTAHGGTNNV 241
DB 1110 EVEIKANGKRADFTETFSSTHGEAPAEWTTIDADGGQGWCLSSGQDMLTAHGGTNNV 1169
QY 242 VASFSWNGMALPNPNLYLSKDVGTGATKVKYVAVNDGFPGDHYAVMI SKGTGNAGDFTTV 301
DB 1170 VASFSWNGMALPNPNLYLSKDVGTGATKVKYVAVNDGFPGDHYAVMI SKGTGNAGDFTTV 1229
QY 302 FEETPENGKNGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYILLD 361
DB 1230 FEETPENGKNGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLYILLD 1289
QY 362 DIQFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
DB 1290 DIQFTMGSPPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1349
QY 422 VCVNVTINPTQFNPNVNL 439
DB 1350 ECVNVTINPTQFNPNVNL 1367

RESULT 8
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; NUMBER OF INVENTIONS: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
```

COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,308A
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/119,361
FILING DATE: 10-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,441
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 21-93C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-308A-10

Query Match 86.2%; Score 2058; DB 3; Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;
Qy 3 PNPENPNPGTTTLESFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICVSS 62
Db 952 PNPENPNPGTTTLESFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICVSS 1008
Qy 63 ASY-INTEGPNPNPNTLTPPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1009 ESFGLGGIGVLTDPNYLITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASNF 1068
Qy 122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLD 181
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQCTDFFWINLD 1126
Qy 182 EVEIKANGKRADFTETTESSTHGEAPAEWTTIDADGQGGWCLSSGQDLWLTAGGTNV 241
Db 1127 EVEIKANGKRADFTETTESSTHGEAPAEWTTIDADGQGGWCLSSGQDLWLTAGGTNV 1186
Qy 242 VASFSWNGMALNPNYLISKDVTGATKYVYAVNDGPGPDHYAVMISKTGTNAGDFTVV 301
Db 1187 VASFSWNGMALNPNYLISKDVTGATKYVYAVNDGPGPDHYAVMISKTGTNAGDFTVV 1246
Qy 302 FEETPNKNGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNSDLNIIYLLD 361
Db 1247 FEETPNKNGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNSDLNIIYLLD 1306
Qy 362 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
Db 1307 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1366
Qy 422 VCVNVTINPTQNPVQNL 439
Db 1367 ECVNVTINPTQNPVQNL 1384

RESULT 9

US-08-822-324-6
; Sequence 6, Application US/08822324
; Patent No. 6129917

GENERAL INFORMATION:
APPLICANT: Potempa, Jan S.
APPLICANT: Travis, James
APPLICANT: Genco, Caroline A.
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
FORPHYOMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
METHODS
TITLE OF INVENTION: FORPHYOMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
METHODS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,945
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 103-95 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 488-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-822-324-6

Query Match 86.2%; Score 2058; DB 3; Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;
Qy 3 PNPENPNPGTTTLESFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICVSS 62
Db 952 PNPENPNPGTTTLESFENGIPASWKTIDADGDNWTTTPPGGTSFAGHNSAICVSS 1008
Qy 63 ASY-INTEGPNPNPNTLTPPELSLPGGGTLTFWVCAQDANYASEHYAVYASSTGNDASNF 121
Db 1009 ESFGLGGIGVLTDPNYLITPALDLPNGGKLTFFWCAQDANYASEHYAVYASSTGNDASNF 1068
Qy 122 ANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLD 181
Db 1069 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQCTDFFWINLD 1126
Qy 182 EVEIKANGKRADFTETTESSTHGEAPAEWTTIDADGQGGWCLSSGQDLWLTAGGTNV 241
Db 1127 EVEIKANGKRADFTETTESSTHGEAPAEWTTIDADGQGGWCLSSGQDLWLTAGGTNV 1186
Qy 242 VASFSWNGMALNPNYLISKDVTGATKYVYAVNDGPGPDHYAVMISKTGTNAGDFTVV 301
Db 1187 VASFSWNGMALNPNYLISKDVTGATKYVYAVNDGPGPDHYAVMISKTGTNAGDFTVV 1246
Qy 302 FEETPNKNGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNSDLNIIYLLD 361
Db 1247 FEETPNKNGGARFGLSTEADGAKPQSVWIERTVLDLPAGTKYVAFRHYNSDLNIIYLLD 1306
Qy 362 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 421
Db 1307 DIQFTMGSGPTDITYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVVEKYTAGVSPK 1366

QY 422 VCVNVTINPTQFNPVONL 439
US-09-490-931-10
; Sequence 10, Application US/09490931
; Patent No. 6274718
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,931
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,308
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-490-931-10
Query Match 86.2%; Score 2058; DB 3; Length 1704;
Best Local Similarity 88.8%; Pred. No. 5.3e-174;
Matches 389; Conservative 12; Mismatches 31; Indels 6; Gaps 4;
QY 3 PNPENPNPGTTTLESFENGIPASWKITDADGDNWTTTPPGTGFAGHNSAICVSS 62
DB 952 PNPENPNPGTTTLESFENGIPASWKITDADGDNWTTTPPGTGFAGHNSAICVSS 1008
QY 63 ASY-INFGPQNDNVLVTPPELSLPGGGLTWCAQDANYASEHYAVYASSTGNDASNF 121
DB 1009 ESFGLGGIGVLTPDNLITPDLNPGGKLTFWCAQDANYASEHYAVYASSTGNDASNF 1068
QY 122 ANALLEEVLTAKTVVTAPEAIRGTRVQSGTWYQKTVQLPAGTKYVAFRHFPGCTDFFWINLD 181
DB 1069 TNALLSEITAKG-VRSEPAIRG-RIQGWTRQKTVDLPGTKYVAFRHFQSTDMFYIDL 1126
QY 182 EVEIKANGKRADFTTFSSSTHGEAPAEWTTTIDADGGQGWCLSSGQLDWLTAHGGTNV 241

DB 1127 EVEIKANGKRADFTTFSSSTHGEAPAEWTTTIDADGGQGWCLSSGQLDWLTAHGGTNV 1186
QY 242 VASFSWNGMALNPNDYLIISKDVTGATKVKYYVAVNDGFPDGHYAVMISKTGTNAGDFTVV 301
DB 1187 VASFSWNGMALNPNDYLIISKDVTGATKVKYYVAVNDGFPDGHYAVMISKTGTNAGDFTVV 1246
QY 302 FEETPNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLD 361
DB 1247 FEETPNGINKGARFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNYILLD 1306
QY 362 DIQTMGSGPTTDTYTVTVRDGKIKIEGLTETTFEEDGVATGHNHEVCVEKYTAGVSPK 421
DB 1307 DIQTMGSGPTTDTYTVTVRDGKIKIEGLTETTFEEDGVATGHNHEVCVEKYTAGVSPK 1366
QY 422 VCVNVTINPTQFNPVONL 439
DB 1367 ECVNVTINPTQFNPVKNL 1384
RESULT 11
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulskie-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Gylvaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1087 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-570-311-8

```
Query Match      86.1%; Score 2055; DB 2; Length 1087;
Best Local Similarity 88.4%; Pred. No. 4.9e-174;
Matches 389; Conservative 13; Mismatches 32; Indels 6; Gaps 4;

QY 1 GTNPENPNPGTTTLLSEFENGIPASWKTIDADGNNWTTTPPGGTSPAGHNSAICV 60
Db 333 GTNPENPNPGTTTLLSEFENGIPASWKTIDADGNNWTTTPPGGTSPAGHNSAICV 389

QY 61 SSASY-INFEQPNPNLYLTPPELSPGGGTLTFWVCAQDANYASEHYAVASSTGNDAS 119
Db 390 YSEFGLGGIGVLTDPNLYITPALDLANGGKLTFWVCAQDANYASEHYAVASSTGNDAS 449

QY 120 NFANALLEEVLTAQVTAPEAIRGTVOGTWYOKTVOLPAGTKYVAFRHFQCTDFWIN 179
Db 450 NFNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQCTDFMEYID 507

QY 180 LDEVEIKANGKRADFTTFESSSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGT 239
Db 508 LDEVEIKANGKRADFTTFESSSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGT 567

QY 240 NVVASFWSNGMALNPDPNLYLSKDVGTATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFT 299
Db 568 NVVASFWSNGMALNPDPNLYLSKDVGTATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFT 627

QY 300 VPEETPNGLKGGARFGLSTEADGAKPOSVWERTVDLPAGTKYVAFRHYNCSDLNLYL 359
Db 628 VPEETPNGLKGGARFGLSTEADGAKPOSVWERTVDLPAGTKYVAFRHYNCSDLNLYL 687

QY 360 LDDIQTMGSPPTDYYTVYRDGTIKKEGLTETTFEEDGVATGNHEYCVVKYTAGVS 419
Db 688 LDDIQTMGSPPTDYYTVYRDGTIKKEGLTETTFEEDGVATGNHEYCVVKYTAGVS 747

QY 420 PKVCVNVTTINPTQFNPVQNL 439
Db 748 PKVCVNVTTINPTQFNPVQNL 767

RESULT 12
US-08-353-485-8
; Sequence 8, Application US/08353485
; Patent No. 5830710
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1087 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-353-485-8

Query Match      86.1%; Score 2055; DB 2; Length 1087;
Best Local Similarity 88.4%; Pred. No. 4.9e-174;
Matches 389; Conservative 13; Mismatches 32; Indels 6; Gaps 4;

QY 1 GTNPENPNPGTTTLLSEFENGIPASWKTIDADGNNWTTTPPGGTSPAGHNSAICV 60
Db 333 GTNPENPNPGTTTLLSEFENGIPASWKTIDADGNNWTTTPPGGTSPAGHNSAICV 389

QY 61 SSASY-INFEQPNPNLYLTPPELSPGGGTLTFWVCAQDANYASEHYAVASSTGNDAS 119
Db 390 YSEFGLGGIGVLTDPNLYITPALDLANGGKLTFWVCAQDANYASEHYAVASSTGNDAS 449

QY 120 NFANALLEEVLTAQVTAPEAIRGTVOGTWYOKTVOLPAGTKYVAFRHFQCTDFWIN 179
Db 450 NFNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVLDPAGTKYVAFRHFQCTDFMEYID 507

QY 180 LDEVEIKANGKRADFTTFESSSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGT 239
Db 508 LDEVEIKANGKRADFTTFESSSTHGEAPAEWTTIDADGGQWMLCLSSGQDMLTAHGGT 567

QY 240 NVVASFWSNGMALNPDPNLYLSKDVGTATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFT 299
Db 568 NVVASFWSNGMALNPDPNLYLSKDVGTATKVKYYAVNDGPPGDHYAVMISKTGTNAGDFT 627

QY 300 VPEETPNGLKGGARFGLSTEADGAKPOSVWERTVDLPAGTKYVAFRHYNCSDLNLYL 359
Db 628 VPEETPNGLKGGARFGLSTEADGAKPOSVWERTVDLPAGTKYVAFRHYNCSDLNLYL 687

QY 360 LDDIQTMGSPPTDYYTVYRDGTIKKEGLTETTFEEDGVATGNHEYCVVKYTAGVS 419
Db 688 LDDIQTMGSPPTDYYTVYRDGTIKKEGLTETTFEEDGVATGNHEYCVVKYTAGVS 747

QY 420 PKVCVNVTTINPTQFNPVQNL 439
Db 748 PKVCVNVTTINPTQFNPVQNL 767

RESULT 13
US-08-570-311-27
; Sequence 27, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
```



```
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UFI5.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-27

Query Match 86.1%; Score 2055; DB 2; Length 1358;
Best Local Similarity 88.4%; Pred. No. 6.9e-174;
Matches 389; Conservative 13; Mismatches 32; Indels 6; Gaps 4;

QY 1 GTPNPENPNPGTTTISEFENGIPASWKTIADGDNWTTTPPGTSGFAGHNSAICV 60
DB 604 GTPNPENPNPGTTTISEFENGIPASWKTIADGDNWTTTPPGTSGFAGHNSAICV 660
QY 61 SSASY-INFEQPNPNPNLYVTPELSLPGGGTTLTFWVCAQDANYASEHYAVYASSTGNDS 119
DB 661 YSEFGLGGIGVLTDPDNYLITPALDLANGKGLTFWVCAQDANYASEHYAVYASSTGNDS 720
QY 120 NFNALLLEVLTAKTVVTAPAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWIN 179
DB 721 NFNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFYID 778
QY 180 LDEVIKANKRADTFETPESSTHGEAPAEWTTIDADGQGWCLSSGQDMLTAHGTTV 239
DB 779 LDEVIKANKRADTFETPESSTHGEAPAEWTTIDADGQGWCLSSGQDMLTAHGTTV 838
QY 240 NVASFWSNGMALPNPNLYLSKDVTKATKVKYKYAVNDGFGPDGHYAVMIKGTGNAGDFT 299
DB 839 NVASFWSNGMALPNPNLYLSKDVTKATKVKYKYAVNDGFGPDGHYAVMIKGTGNAGDFT 898
QY 300 VVFEETPNGINKGARGFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNLYIL 359
DB 899 VVFEETPNGINKGARGFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNLYIL 958
QY 360 LDDIQFTMGSGPTDPTDYTVYVRDGTIKKGLTETTFEEDGVATGNHEYCVEVKYTAGVS 419
DB 959 LDDIQFTMGSGPTDPTDYTVYVRDGTIKKGLTETTFEEDGVATGNHEYCVEVKYTAGVS 1018
QY 420 PKVCNVNTINPTQNPVNQL 439
DB 1019 PKVCNVNTINPTQNPVNQL 1038
```

```
RESULT 14
US-09-066-330-10
; Sequence 10, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCT/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-10
```

```
Query Match 85.8%; Score 2048; DB 4; Length 1706;
Best Local Similarity 88.1%; Pred. No. 4.1e-173;
Matches 386; Conservative 14; Mismatches 32; Indels 6; Gaps 4;

QY 3 PNPENPNPGTTTISEFENGIPASWKTIADGDNWTTTPPGTSGFAGHNSAICVSS 62
DB 954 PNPENPNPGTTTISEFENGIPASWKTIADGDNWTTTPPGTSGFAGHNSAICVSS 1010
QY 63 ASY-INFEQPNPNPNLYVTPELSLPGGGTTLTFWVCAQDANYASEHYAVYASSTGNDSNF 121
DB 1011 ESFGLGGIGVLTDPDNYLITPALDLNGKGLTFWVCAQDANYASEHYAVYASSTGNDSNF 1070
QY 122 ANALLEVLTAKTVVTAPAIRTRVQGTWYQKTVQLPAGTKYVAFRHFQCTDFFWINLD 181
DB 1071 TNALLEETITAKG-VRSPEAIRG-RIQGTWRQKTVDLPAQTKYVAFRHFQSTDMFYIDLD 1128
QY 182 EVEIKANKRADTFETPESSTHGEAPAEWTTIDADGQGWCLSSGQDMLTAHGTTV 241
DB 1129 EVEIKANKRADTFETPESSTHGEAPAEWTTIDADGQGWCLSSGQDMLTAHGTTV 1188
QY 242 VASFWSNGMALPNPNLYLSKDVTKATKVKYKYAVNDGFGPDGHYAVMIKGTGNAGDFTV 301
DB 1189 VASFWSNGMALPNPNLYLSKDVTKATKVKYKYAVNDGFGPDGHYAVMIKGTGNAGDFTV 1248
QY 302 FEETPNGINKGARGFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNLYILD 361
DB 1249 FEETPNGINKGARGFGLSTEADGAKPQSVMIERTVDLPAGTKYVAFRHYNCSDLNLYILD 1308
QY 362 DIQFTMGSGPTDPTDYTVYVRDGTIKKGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 421
DB 1309 DIQFTMGSGPTDPTDYTVYVRDGTIKKGLTETTFEEDGVATGNHEYCVEVKYTAGVSPK 1368
QY 422 VCVNTINPTQNPVNQL 439
DB 1369 KCVNTVNSTQNPVNQL 1386
```

```
RESULT 15
US-09-066-330-11
; Sequence 11, Application US/09066330A
; Patent No. 6511666
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
```

```
; CURRENT APPLICATION NUMBER: US/09/066,330A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: PN 6275
; EARLIER FILING DATE: 1995-10-30
; EARLIER APPLICATION NUMBER: PCI/AU96/00673
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-066-330-11

Query Match      85.0%; Score 2030.5; DB 4; Length 1732;
Best Local Similarity 87.3%; Pred No. 1.5e-171;
Matches 385; Conservative 15; Mismatches 34; Indels 7; Gaps 5;

QY 1 GTPNPNPNPFGT-TTISESFENGIPASWKTIDADGDNNTTTPPPGGTSFAGHNSAIC 59
Db 969 GTPNPNPNPNPCTTISESFENGIPASWKTIDADGDNNTTTPPPGGTSFAGHNSAIC 1025
QY 60 VSSASY-INTEGPNPNPNLVTPELSLPGGTLTFWVCAQDANYASEHYAYASSTGND 118
Db 1026 VYSEFGLGGIGVLTDPNLTLPALDLPNGGKLTFWVCAQDANYASEHYAYASSTGND 1085
QY 119 SNFANALLEEVLTAKTVTVAPEAIRGTRVOGTWYOKTVOLPAGTKYVAFRHFPGCTDPFWI 178
Db 1086 SNFTNALLEETITAKG-VRSKPAIRG-RIQGTWQKTVDLDPAGTKYVAFRHFQSTDMFYI 1143
QY 179 NLDEVEIKANGKRADFTETPESSTHGEAPAEWTTIDADGGQGWMLCLSSGOLDWLTAGG 238
Db 1144 DLDEVEIKANGKRADFTETPESSTHGEAPAEWTTIDADGGQGWMLCLSSGOLDWLTAGG 1203
QY 239 TNVVASFSWNGMALNPDNLYISKDVTGATKYKYAVNDGPPGDHYAVMISKTGTNAGDF 298
Db 1204 SNVVSFSWNGMALNPDNLYISKDVTGATKYKYAVNDGPPGDHYAVMISKTGTNAGDF 1263
QY 299 TVVFEETPNGINKGARFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 358
Db 1264 TVVFEETPNGINKGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYI 1323
QY 359 LLDDIQFTMGGSPTFTDYTYTVYRDGTKIKSGLTETTFEEDGVATGNHEYCVVKYTAGV 418
Db 1324 LLDDIQFTMGGSPTFTDYTYTVYRDGTKIKSGLTETTFEEDGVATGNHEYCVVKYTAGV 1383
QY 419 SPKVCVNVNTINPTQFNPVQNL 439
Db 1384 SPKKCVNVNTVNSTQFNPVQNL 1404
```

Search completed: May 18, 2004, 11:49:17
Job time : 14.5635 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 18, 2004, 11:42:50 ; Search time 34.0043 Seconds
(without alignments)
3592.387 Million cell updates/sec

Title: US-08-570-311-22

Perfect score: 2388

Sequence: 1 GTPNPNPNGTTTSLBSF.....PKVCNVNTPNQPNVQNL 439

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|------------------|
| 1 | 2048 | 85.8 | 1706 | 14 | US-10-229-066-10 |
| 2 | 2030.5 | 85.0 | 1732 | 14 | US-10-229-066-11 |
| 3 | 727 | 30.4 | 419 | 15 | US-10-174-695-3 |
| 4 | 711.5 | 29.8 | 419 | 15 | US-10-174-695-5 |
| 5 | 303 | 12.7 | 231 | 15 | US-10-174-695-6 |
| 6 | 233.5 | 9.8 | 196 | 15 | US-10-174-695-4 |
| 7 | 153 | 6.4 | 29 | 15 | US-10-387-977-18 |
| 8 | 153 | 6.4 | 29 | 15 | US-10-387-977-21 |
| 9 | 153 | 6.4 | 29 | 15 | US-10-387-977-23 |
| 10 | 153 | 6.4 | 29 | 15 | US-10-387-977-24 |
| 11 | 153 | 6.4 | 29 | 15 | US-10-387-977-26 |
| 12 | 147 | 6.2 | 29 | 15 | US-10-387-977-17 |
| 13 | 145 | 6.1 | 29 | 15 | US-10-387-977-25 |
| 14 | 137 | 5.7 | 25 | 15 | US-10-387-977-81 |
| 15 | 136 | 5.7 | 29 | 15 | US-10-387-977-20 |

| | | | | | | |
|----|-------|-----|------|----|----------------------|-------------------|
| 16 | 133 | 5.6 | 872 | 12 | US-10-282-122A-55467 | Sequence 55467, A |
| 17 | 132.5 | 5.5 | 555 | 14 | US-10-156-761-14592 | Sequence 14592, A |
| 18 | 132 | 5.5 | 25 | 15 | US-10-387-977-83 | Sequence 83, Appl |
| 19 | 132 | 5.5 | 25 | 15 | US-10-387-977-84 | Sequence 84, Appl |
| 20 | 132 | 5.5 | 25 | 15 | US-10-387-977-85 | Sequence 85, Appl |
| 21 | 132 | 5.5 | 960 | 14 | US-10-311-879-15 | Sequence 15, Appl |
| 22 | 131.5 | 5.5 | 1638 | 12 | US-10-206-576-258 | Sequence 258, App |
| 23 | 131.5 | 5.5 | 1638 | 12 | US-10-206-576-262 | Sequence 262, App |
| 24 | 131.5 | 5.5 | 1638 | 12 | US-10-206-576-266 | Sequence 266, App |
| 25 | 131.5 | 5.5 | 1728 | 12 | US-10-282-122A-56997 | Sequence 56997, A |
| 26 | 131 | 5.5 | 2435 | 12 | US-10-282-122A-47453 | Sequence 47453, A |
| 27 | 130 | 5.4 | 25 | 15 | US-10-387-977-77 | Sequence 77, Appl |
| 28 | 130 | 5.4 | 25 | 15 | US-10-387-977-80 | Sequence 80, Appl |
| 29 | 129 | 5.4 | 24 | 14 | US-10-229-066-6 | Sequence 6, Appl |
| 30 | 129 | 5.4 | 691 | 15 | US-10-369-493-23473 | Sequence 23473, A |
| 31 | 125.5 | 5.3 | 669 | 12 | US-10-206-576-264 | Sequence 264, App |
| 32 | 125 | 5.2 | 25 | 15 | US-10-387-977-76 | Sequence 76, Appl |
| 33 | 125 | 5.2 | 25 | 15 | US-10-387-977-79 | Sequence 79, Appl |
| 34 | 123.5 | 5.2 | 1095 | 14 | US-10-242-056-34 | Sequence 34, Appl |
| 35 | 123.5 | 5.2 | 1095 | 15 | US-10-282-794A-34 | Sequence 34, Appl |
| 36 | 121 | 5.1 | 1649 | 15 | US-10-369-493-18460 | Sequence 18460, A |
| 37 | 120 | 5.0 | 1833 | 14 | US-10-175-282-4 | Sequence 4, Appl |
| 38 | 120 | 5.0 | 1833 | 14 | US-10-175-275-4 | Sequence 4, Appl |
| 39 | 120 | 5.0 | 1992 | 14 | US-10-175-282-3 | Sequence 3, Appl |
| 40 | 120 | 5.0 | 1992 | 14 | US-10-175-275-3 | Sequence 3, Appl |
| 41 | 119 | 5.0 | 1333 | 12 | US-10-282-122A-5546 | Sequence 5546, A |
| 42 | 118.5 | 5.0 | 2122 | 9 | US-09-813-214A-9 | Sequence 9, Appl |
| 43 | 118 | 4.9 | 2353 | 9 | US-09-797-862-33 | Sequence 33, Appl |
| 44 | 117 | 4.9 | 22 | 14 | US-10-229-066-4 | Sequence 4, Appl |
| 45 | 117 | 4.9 | 29 | 15 | US-10-387-977-19 | Sequence 19, Appl |

ALIGNMENTS

RESULT 1

US-10-229-066-10 ; Sequence 10, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Rhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-10

Query Match 85.8%; Score 2048; DB 14; Length 1706;
Best Local Similarity 88.1%; Pred. No. 1e-177;
Matches 386; Conservative 14; Mismatches 32; Indels 6; Gaps 4;

| | | |
|----|------|--|
| QY | 3 | PNPNPNGTTTSLBSFENGIPASWKTIADGDGNWTTTTPPGTSPAGHNSAICVSS 62 |
| Db | 954 | PNPNPNGTTTSLBSFENGIPASWKTIADGDGHGKPGNAPG---TAGYNSNCVYS 1010 |
| QY | 63 | ASY-INFEQPNPDNYLVTPFELSPGGGTLTFWVCAQADANYASEHYAVVASTGNDASNF 121 |
| Db | 1011 | ESFGLGGIGVLPDNYLVTPFELSPGGGTLTFWVCAQADANYASEHYAVVASTGNDASNF 1070 |

122 ANALLEEVLTAKTVTAPALRGTRVOGTWYOKTVQVLPAGTKYVAFRHFHGTDFWNLND 181
1071 TNALLEETITAKG-VRSPPEARWG-RIQGTWRQKTVDLPAGTKYVAFRHFHGTDFWNLND 1128
182 EVEIKANGKRAADFTETPESSTHGEAPAEWTTTIDADGGQGWLCCLSSGQLDWTAGHGTNV 241
1129 EVEIKANGKRAADFTETPESSTHGEAPAEWTTTIDADGGQGWLCCLSSGQLDWTAGHGTNV 1188
242 VASFSWNGMALPNPNYLISKDVTGATKVKYKYYAVNDGPGDGHYAVMISKTGTNAGDFTVV 301
1189 VSSFSWNGMALPNPNYLISKDVTGATKVKYKYYAVNDGPGDGHYAVMISKTGTNAGDFTVV 1248
302 FEETPENGINKGARFGLSTEADGAKPOSWIERVTDLPAGTKYVAFRHYNCSDNLYILLD 361
1249 FEETPENGINKGARFGLSTEADGAKPOSWIERVTDLPAGTKYVAFRHYNCSDNLYILLD 1308
362 DIQFTMGSSPTDITYTYVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKKTAGVSPK 421
1309 DIQFTMGSSPTDITYTYVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKKTAGVSPK 1368
422 VCVNVTINPTQFNPQNL 439
1369 KCVNVTINPTQFNPQNL 1386

RESULT 2
US-10-229-066-11
; Sequence 11, Application US/10229066
; Publication No. US20030157637A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric C.
; APPLICANT: Bhogal, Peter S.
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: DIAGNOSTICS AND TREATMENTS OF PERIODONTAL DISEASE
; FILE REFERENCE: Reynolds
; CURRENT APPLICATION NUMBER: US/10/229,066
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/066,330
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: PN 6275
; PRIOR FILING DATE: 1995-10-30
; PRIOR APPLICATION NUMBER: PCT/AU96/00673
; PRIOR FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-229-066-11

Query Match 85.0%; Score 2030.5; DB 14; Length 1732;
Best Local Similarity 87.3%; Pred. No. 4.3e-176; Indels 7; Gaps 5;
Matches 385; Conservative 15; Mismatches 34; Indels 7; Gaps 5;

1 GTPNPNPNPCT-TTLESFENGIPASWKITIDADGGNNWTTTPPPGTSFAGHNSAIC 59
969 GTPNPNPNPNPCT-TTLESFENGIPASWKITIDADGGHGWKPGNAPG---IAGYNSGNC 1025
60 VSSASY-INFEQPONPNYLVTPELSLPGGGTLTFWVCAQADANYASEHYAVASSTGND 118
1026 VYSEFGLGGIGLVLPDNLITPALDLPNGGKLTFWVCAQADANYASEHYAVASSTGND 1085
119 SNFANALLEEVLTAKTVTAPALRGTRVOGTWYOKTVQVLPAGTKYVAFRHFHGTDFWNL 178
1086 SNFTALLEETITAKG-VRSPPEARWG-RIQGTWRQKTVDLPAGTKYVAFRHFHGTDFWNL 1143
179 NLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTTIDADGGQGWLCCLSSGQLDWTAGH 238
1144 DLDEVEIKANGKRAADFTETPESSTHGEAPAEWTTTIDADGGQGWLCCLSSGQLDWTAGH 1203
239 TNVASFWSWNGMALPNPNYLISKDVTGATKVKYKYYAVNDGPGDGHYAVMISKTGTNAGD 298

1204 SNVVSFSWNGMALPNPNYLISKDVTGATKVKYKYYAVNDGPGDGHYAVMISKTGTNAGD 1263
299 TVVFEETPENGINKGARFGLSTEADGAKPOSWIERVTDLPAGTKYVAFRHYNCSDNLYI 358
1264 TVVFEETPENGINKGARFGLSTEADGAKPOSWIERVTDLPAGTKYVAFRHYNCSDNLYI 1323
359 LLDDIQTFTMGSSPTDITYTYVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKKTAGV 418
1324 LLDDIQTFTMGSSPTDITYTYVYRDGTKIKEGLTETTFEEDGVATGNHGYCDEVKKTAGV 1383
419 SPKVCNVTINPTQFNPQNL 439
1384 SPKVCNVTINPTQFNPQNL 1404

RESULT 3
US-10-174-695-3
; Sequence 3, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695-3

Query Match 30.4%; Score 727; DB 15; Length 419;
Best Local Similarity 75.8%; Pred. No. 1.1e-57; Indels 6; Gaps 4;
Matches 144; Conservative 10; Mismatches 30; Indels 6; Gaps 4;

3 PNPNNPNPCT-TTLESFENGIPASWKITIDADGGNNWTTTPPPGTSFAGHNSAICVSS 62
235 PNPNNPNPCT-TTLESFENGIPASWKITIDADGGHGWKPGNAPG---IAGYNSGNCVSS 291
63 ASY-INFEQPONPNYLVTPELSLPGGGTLTFWVCAQADANYASEHYAVASSTGNDASNF 121
292 ESFGLGGIGLVLPDNLITPALDLPNGGKLTFWVCAQADANYASEHYAVASSTGNDASNF 351
122 ANALLEEVLTAKTVTAPALRGTRVOGTWYOKTVQVLPAGTKYVAFRHFHGTDFWNLND 181
352 TNALLEETITAKG-VRSPPEARWG-RIQGTWRQKTVDLPAGTKYVAFRHFHGTDFWNLND 409
182 EVEIKANGKR 191
410 EVEIKANGKR 419

RESULT 4
US-10-174-695-5
; Sequence 5, Application US/10174695
; Publication No. US20030232022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Slakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Barr, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 52928200700
; CURRENT APPLICATION NUMBER: US/10/174,695

```

Db      92 RYDDTFE--AGKKYTMTMRAGMGDTDEVED-----DSPASYTYTVYRDGTLK 141

Qy      389 EGLTETTFEEBDGVATGNHXYCVEVKYTAGVSPKVCVANVTI-NP*QFNVPQNL 439
          |||||
          |||||FEEDGVAAGNHXYCVEVKYTAGVSPKCKDVTVEGSNEFAPQNL 193
          |||||

RESULT 6
US-10-174-695--A
; Sequence 4, Application US/10174695
; Publication No. US2003032022A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: Stakeski, Nada
; APPLICANT: Chen, Chao Guang
; APPLICANT: Baiz, Ian George
; TITLE OF INVENTION: P. GINGIVALIS ANTIGENIC COMPOSITION
; FILE REFERENCE: 529282000700
; CURRENT APPLICATION NUMBER: US/10/174,695
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: PCT/AU00/01588
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: AU PQ 4859
; PRIOR FILING DATE: 1999-12-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: fastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-174-695--A

```

```

QY      3  PNPNNPNEGTTTSESFENGIPASWKITDADGDNWNTTTTTPPGTSPAGHNSAICV 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      44  PNPNNPNEGTTTSESFENGIPASWKITDADGDNWNTTTTTPPGTSPAGHNSAICV 98
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-10-387-977-18
; Sequence 18, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeski, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: Fasteq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-18

```

```

Query Match          6.4%; Score 153; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 7.3e-07;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
 |||||
 Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 8

US-10-387-977-21
 ; Sequence 21, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 21
 ; LENGTH: 29
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-21

Query Match 6.4%; Score 153; DB 15; Length 29;
 Best Local Similarity 100.0%; Pred. No. 7.3e-07;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
 |||||
 Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 9

US-10-387-977-23
 ; Sequence 23, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 23
 ; LENGTH: 29
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-23

Query Match 6.4%; Score 153; DB 15; Length 29;
 Best Local Similarity 100.0%; Pred. No. 7.3e-07;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
 |||||
 Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 10

US-10-387-977-24
 ; Sequence 24, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 24
 ; LENGTH: 29
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-24

Query Match 6.4%; Score 153; DB 15; Length 29;
 Best Local Similarity 100.0%; Pred. No. 7.3e-07;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 DYTIVYRDGTKIKEGLTETTFEEDGVAT 403
 |||||
 Db 1 DYTIVYRDGTKIKEGLTETTFEEDGVAT 29

RESULT 11

US-10-387-977-26
 ; Sequence 26, Application US/10387977
 ; Publication No. US20040005276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reynolds, Eric Charles
 ; APPLICANT: O'Brien-Simpson, Neil Martin
 ; APPLICANT: Slakeski, Nada
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
 ; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
 ; FILE REFERENCE: 52928200301
 ; CURRENT APPLICATION NUMBER: US/10/387,977
 ; CURRENT FILING DATE: 2003-07-18
 ; PRIOR APPLICATION NUMBER: US 09/423,056
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00311
 ; PRIOR FILING DATE: 1998-04-30
 ; PRIOR APPLICATION NUMBER: AU PO 6528
 ; PRIOR FILING DATE: 1997-04-30
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 26
 ; LENGTH: 29
 ; TYPE: PRT
 ; ORGANISM: Porphyromonas gingivalis
 US-10-387-977-26

Query Match 6.4%; Score 153; DB 15; Length 29;
 Best Local Similarity 100.0%; Pred. No. 7.3e-07;

Query Match 6.1%; Score 145; DB 15; Length 29;

RESULT 15
US-10-387-977-20
; Sequence 20, Application US/10387977
; Publication No. US20040005276A1
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Eric Charles
; APPLICANT: O'Brien-Simpson, Neil Martin
; APPLICANT: Slakeeki, Nada
; TITLE OF INVENTION: SYNTHETIC PEPTIDE CONSTRUCTS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PERIODONTITIS ASSOCIATED WITH
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS
; FILE REFERENCE: 529282000301
; CURRENT APPLICATION NUMBER: US/10/387,977
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/423,056
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: PCT/AU98/00311
; PRIOR FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: AU PO 6528
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-387-977-20

